PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/47, 16/18, C12N 15/62 — A2

(11) International Publication Number: WO 99/03993

(43) International Publication Date: 28 January 1999 (28.01.99)

(21) International Application Number: PCT/US98/14544

(22) International Filing Date: 17 July 1998 (17.07.98)

(30) Priority Data:

60/053,153	18 July 1997 (18.07.97)	US
60/053,244	18 July 1997 (18.07.97)	US
60/055,804	15 August 1997 (15.08.97)	US
60/055,853	15 August 1997 (15.08.97)	US

(71) Applicant: SCHERING CORPORATION [US/US]; 2000 Galloping Hill Road, Kenilworth, NJ 07033 (US).

(72) Inventor: JOHNSON, James, A.; 205 Mary Alice Drive, Los Gatos, CA 95032 (US).

(74) Agents: MCLAUGHLIN, Jaye, P. et al.; Schering-Plough Corporation, Patent Dept., K-6-1 1990, 2000 Galloping Hill Road, Kenilworth, NJ 07033-0530 (US).

(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CZ, EE, GE, HR, HU, ID, IL, IS, JP, KG, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UZ, VN, YU, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: SUPPRESSORS OF CYTOKINE SIGNALING; RELATED REAGENTS

(57) Abstract

Purified genes encoding intracellular regulatory molecules from a human, reagents related thereto including purified proteins, specific antibodies, and nucleic acids encoding these molecules are provided. Methods of using said reagents and diagnostic kits are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SUPPRESSORS OF CYTOKINE SIGNALING; RELATED REAGENTS

This filing is a PCT Application claiming priority to provisional U.S. Patent Applications USSN 60/055,804, filed August 15, 1997, and USSN 60/053,153, filed July 18, 1997. Also incorporated by reference are provisional U.S. Patent Applications USSN 60/055,853, filed August 15, 1997, and USSN 60/053,244, filed July 18, 1997.

FIELD OF THE INVENTION

The present invention pertains to compositions related to proteins which function, e.g., in suppressing intracellular signaling pathways, e.g., cytokine signaling. In particular, it provides purified genes, proteins, antibodies, and related reagents useful, e.g., to regulate growth hormone-like or cytokine-regulated intracellular processes, including transcription or genes in various cell types, including immune cells.

BACKGROUND OF THE INVENTION

Recombinant DNA technology refers generally to the technique of integrating genetic information from a donor source into vectors for subsequent processing, such as through introduction into a host, whereby the transferred genetic information is copied and/or expressed in the new environment. Commonly, the genetic information exists in the form of complementary DNA (cDNA) derived from messenger RNA (mRNA) coding for a desired protein product. The carrier is frequently a plasmid having the capacity to incorporate cDNA for later replication in a host and, in some cases, actually to control expression of the cDNA and thereby direct synthesis of the encoded product in the host.

For some time, it has been known that the mammalian immune response is based on a series of complex cellular interactions, called the "immune network". Recent research has provided new insights into the inner

workings of this network. While it remains clear that much of the response does, in fact, revolve around the network-like interactions of lymphocytes, macrophages, granulocytes, and other cells, immunologists now generally hold the opinion that soluble proteins, known as lymphokines, cytokines, or monokines, play a critical role in controlling these cellular interactions. Thus, there is considerable interest in the isolation, characterization, and mechanisms of action of cell modulatory factors, an understanding of which will lead to significant advancements in the diagnosis and therapy of numerous medical abnormalities, e.g., immune system disorders. Some of these factors are hematopoietic growth factors, e.g., granulocyte colony stimulating factor (G-CSF), and others are regulatory molecules. See, e.g., Thomson (1994; ed.) The Cytokine Handbook (2d ed.) Academic Press, San Diego; Metcalf and Nicola (1995) The Hematopoietic Colony Stimulating Factors Cambridge University Press; and Aggarwal and Gutterman (1991) Human Cytokines Blackwell Pub.

Lymphokines apparently mediate cellular activities in a variety of ways. They have been shown to support the proliferation, growth, and differentiation of, e.g., pluripotential hematopoietic stem cells into vast numbers of progenitors comprising diverse cellular lineages making up a complex immune system. Proper and balanced interactions between cellular components are necessary for a healthy developmental or immune response. The different cellular lineages often respond in a different manner when lymphokines are administered in conjunction with other agents.

In the immune system, many of the effects of known cytokines on gene transcription are known to be mediated by cytokine inducible DNA binding proteins. See, e.g., Paul (ed. 1994) Fundamental Immunology, 3rd ed., Raven Press, New York, NY. The mechanisms of signal transduction have been an area of active recent study, and involve protein phosphorylation and dephosphorylation with, e.g., the Janus kinases (JAKs) and Signal

Transducers and Activators of Transcription (Stats). See, e.g., Ihle (1996) Cell 84:331-334; ;Ivashkiv (1995) Immunity 3:1-4; and Ihle and Kerr (1995) Trends in Genetics 11:69-74.

The lack of knowledge regarding the mechanisms of signaling involved in the regulation of cell cycle or transcriptional elements has hampered the ability of medical science to specifically regulate cell division or cellular responses, including immune responses. The present invention provides compositions which will be important in such regulation.

SUMMARY OF THE INVENTION

The present invention is based in part upon the discovery of intracellular regulatory molecules which can block signal transduction, e.g., through growth factor- or cytokine-receptor superfamily signaling mechanisms. These proteins exhibit a structural feature designated a SOCS box. See Hilton, et al. (1998) Proc. Nat'l Acad. Sci. USA 95:114-119. Moreover, the SOCS3 protein can block the IL-2 induced signaling via the STAT5, establishing function of the SOCS proteins as suppressors of cytokine signaling.

The invention provides a substantially pure or recombinant SOCS14 protein or peptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 2 or 6; a natural sequence SOCS14 of SEQ ID NO: 2 or 6; a fusion protein comprising SOCS14 sequence; a substantially pure or recombinant SOCS15 (also designated WDS11) protein or peptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 4 or 8; a natural sequence SOCS15 (WDS11) of SEQ ID NO: 4 or 8; a fusion protein comprising SOCS15 (WDS11) sequence; a substantially pure or recombinant SOCS17 protein or peptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 10; a natural sequence SOCS17 of SEQ ID NO: 10; a fusion protein comprising SOCS17 sequence; a substantially pure or recombinant SOCS18 protein or peptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID

NO: 12; a natural sequence SOCS18 of SEQ ID NO: 12; a fusion protein comprising SOCS18 sequence; a substantially pure or recombinant SOCS19 protein or peptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 14; a natural sequence SOCS19 of SEQ ID NO: 14; a fusion protein comprising SOCS19 sequence; or a substantially pure or recombinant WDS12 protein or peptide exhibiting identity over a length of at least about 12 amino acids to SEQ ID NO: 16; a natural sequence WDS12 of SEQ ID NO: 16; or a fusion protein comprising WDS12 sequence. In preferred embodiments, the portion is at least about 25 amino acids. In other embodiments, the: SOCS14 comprises a mature sequence of SEQ ID NO: 2 or 6; SOCS15 (WDS11) comprises a mature sequence of SEQ ID NO: 4 or 8; SOCS17 comprises a mature sequence of SEQ ID NO: 10; SOCS18 comprises a mature sequence of SEQ ID NO: 12; SOCS19 comprises a mature sequence of SEQ ID NO. 14; WDS12 comprises a mature sequence of SEQ ID NO: 16; protein or peptide: is from a warm blooded animal selected from a mammal, including a primate; comprises at least one polypeptide segment of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, or 16; exhibits a plurality of portions exhibiting the identity; is a natural allelic variant of SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12; has a length at least about 30 amino acids; exhibits at least two non-overlapping epitopes which are specific for a mammalian SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12; exhibits identity over a length of at least about 20 amino acids to SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12; exhibits at least two non-overlapping epitopes which are specific for a SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12; exhibits identity over a length of at least about 25 amino acids to a primate SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12; is glycosylated; is a synthetic polypeptide; is attached to a solid substrate; is conjugated to another chemical moiety; is a 5-fold or less substitution from natural sequence; or is a deletion or insertion variant from a natural sequence. Various preferred embodiments include

a composition comprising: a sterile SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein or peptide; the SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein or peptide and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration. The invention further provides a fusion protein, comprising: mature protein comprising sequence of SEQ ID NO: 2, 6, 4, 8, 10, 12, 14 or 16; a detection or purification tag, including a FLAG, His6, or Ig sequence; or sequence of another SOCS or WDS protein.

These reagents also make available a kit comprising such a protein or polypeptide, and: a compartment comprising the protein or polypeptide; and/or instructions for use or disposal of reagents in the kit.

Providing an antigen, the invention further provides a binding compound comprising an antigen binding portion from an antibody, which specifically binds to a natural SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein, wherein: the protein is a primate protein; the binding compound is an Fv, Fab, or Fab2 fragment; the binding compound is conjugated to another chemical moiety; or the antibody: is raised against a peptide sequence of a mature polypeptide comprising sequence of SEQ ID NO: 2, 6, 4, 8, 10, 12, 14 or 16; is raised against a mature SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12; is raised to a purified SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12; is immunoselected; is a polyclonal antibody; binds to a denatured SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12; exhibits a Kd to antigen of at least 30 μM; is attached to a solid substrate, including a bead or plastic membrane; is in a sterile composition; or is detectably labeled, including a radioactive or fluorescent label. Preferred kits include those containing the binding compound, and: a compartment comprising the binding compound; and/or instructions for use or disposal of reagents in the kit. Many of the kits

will be used for making a qualitative or quantitative analysis.

Other preferred compositions will be those comprising: a sterile binding compound, or the binding compound and a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration.

The present invention further provides an isolated or recombinant nucleic acid encoding a protein or peptide or fusion protein described above, wherein: the SOCS or WDS family protein is from a mammal, including a primate; or the nucleic acid: encodes an antigenic peptide sequence of SEQ ID NO: 2, 6, 4, 8, 10, 12, 14 or 16; encodes a plurality of antigenic peptide sequences of SEQ ID NO: 2, 6, 4, 8, 10, 12, 14 or 16; exhibits identity to a natural cDNA encoding the segment; is an expression vector; further comprises an origin of replication; is from a natural source; comprises a detectable label; comprises synthetic nucleotide sequence; is less than 6 kb, preferably less than 3 kb; is from a mammal, including a primate; comprises a natural full length coding sequence; is a hybridization probe for a gene encoding the SOCS or WDS family protein; or is a PCR primer, PCR product, or mutagenesis primer. In certain embodiments, the invention provides a cell or tissue comprising such a recombinant nucleic acid. Preferred cells include: a prokaryotic cell; a eukaryotic cell; a bacterial cell; a yeast cell; an insect cell; a mammalian cell; a mouse cell; a primate cell; or a human cell.

Other kit embodiments include a kit comprising the described nucleic acid, and: a compartment comprising the nucleic acid; a compartment further comprising a SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein or polypeptide; and/or instructions for use or disposal of reagents in the kit. In many versions, the kit is capable of making a qualitative or quantitative analysis.

Other nucleic acid embodiments include those which: hybridize under wash conditions of 50° C and less than 500 mM salt to SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, or 15;

exhibits identity over a stretch of at least about 30 nucleotides to a SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12. In other embodiments: the wash conditions are at 55° C and/or 300 mM salt; 60° C and/or 150 mM salt; the identity is over a stretch is at least 55 or 75 nucleotides.

In other embodiments, the invention provides a method of modulating physiology or development of a cell or tissue culture cells comprising introducing into such cell an agonist or antagonist of a SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. General

It is to be understood that this invention is not limited to the particular compositions, methods, and techniques described herein, as such compositions, methods, and techniques may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which is only limited by the appended claims.

As used herein, including the appended claims, singular forms of words such as "a," "an," and "the" include their corresponding plural referents unless the context clearly dictates otherwise. Thus, e.g., reference to "a polynucleotide" includes one or more different polynucleotides, reference to "a composition" includes one or more of such compositions, and reference to "a method" include reference to equivalent steps and methods known to a person of ordinary skill in the art, and so forth.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by a person of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are

described below. All publications, patent applications, patents, and other references discussed above are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the invention is not entitled to antedate any such disclosure by virtue of its prior invention. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety including all figures, references, and drawings.

The proliferation, differentiation, and physiological responses of many cell lineages are regulated by secreted proteins, e.g., cytokines. These molecules often exert their biological effects through binding to cell surface receptors that are associated with one or more members of the Janus Kinase (Jak) family of cytoplasmic tyrosine kinases. For example, cytokine induced receptor dimerization leads to the activation of JAKs, rapid tyrosine phosphorylation of cytoplasmic domains, and subsequent recruitment of various signaling proteins to the receptor complex, including members of the STAT family of transcription factors. The JAK and STAT proteins are enzymes which act to transduce a signal from the cell surface to the nucleus, thereby serving as the pathway to signal the cell to respond physiologically to an external signal. These pathways have been shown to involve certain protein phosphorylation or dephosphorylation steps, thereby leading to response or lack of response by the cell. See, e.g., Ihle (1996) Cell 84:331-334; Ivashkiv (1995) Immunity 3:1-4; Ihle, et al. (1995) Ann. Rev. Immunol. 13:369-398; Ihle and Kerr (1995) Trends in Genetics 11:69-74; and Darnell, et al. (1994) Science 264:1415-1421.

A number of novel genes have been identified from mouse or humans which appear to inhibit STAT function. See, e.g., Yoshimura, et al. (1995) EMBO J. 14:2816-2826; Matsumoto, et al. (1997) Blood 89:3148-3154; Starr, et al. (1997) Nature 387:917-921; Endo, et al. (1997) Nature 387:921-924; and Naka, et al. Nature 387:924-929. The

present invention provides additional genes with sequence related to those, designated Suppressors Of Cytokine Signaling or WDS: SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12.

A primate, e.g., human, SOCS14 cDNA fragment and corresponding open reading frame are provided in (SEQ ID NO: 1 and 2). The translation exhibits significant matching and similarity to other identified SOCS family members. The internal stop codon indicates some errors in the sequence at or near those positions.Additional refined sequence of primate, e.g., human, SOCS14 is provided in SEQ ID NO: 5 and 6.

A rodent, e.g., mouse, SOCS15 cDNA fragment and corresponding open reading frame are provided in SEQ ID NO: 3 and 4. The translation exhibits significant matching and similarity to other identified SOCS family members. The internal stop codon indicates some errors in the sequence at or near those positions.

A rodent, e.g., murine SOCS17 CDNA and corresponding open reading frame are provided in SEQ ID NO: 9 and 10. Nucleotide may be A, C, T, or G at positions: 1680, 1691, 1696, 1704, 1707, 1728, 1740, 1743, 1746, 1755, 1760, 1770, 1773, 1802, 1816, 1817, 1823, 1826, 1827, 1846, 1851, 1857, 1861, 1880, 1885, 1909, 1917, 1920, 1929, 1946, 1953, 1967, 1968, 1980, 1991, 1995, 2001, 2004, 2021, 2033, 2034, 2035, 2036, 2037, 2039, 2040, 2042, 2048, 2051, 2054, 2061, 2075, 2081, 2083, 2084, 2085, 2088, 2105, 2121, 2124, 2132, 2137, 2147, 2149, 2151, 2152, 2160, 2165, 2177, 2179 and 2196; nucleotide may be A or C at position 494; nucleotide may be C or T at positions: 498, 501, 1455, 1524, 1527, 1621, 1829, and 2072; nucleotide may be G or C at positions: 499, 1618, and 1664; nucleotide may be G or T at position 1673; and nucleotide may be A, C, or G at positions: 1819, 1840, and 2089 (see SEQ ID NO: 26).

A primate, e.g., human, SOCS18 nucleotide and corresponding amino acid sequence are provided in SEQ ID NO: 11 and 12. Nucleotide may be A or C at positions: 740, 797, 2139, and 2184; nucleotide may be G or T at positions: 761, 1313, 1508, and 2226; nucleotide may be C

or T at positions 746, 1460, 1499, 2009, 2010, 2199, and 2225; nucleotide may be A or G at positions 788, 863, 1550, 2178, 2188, 2197, and 2211; nucleotide may be G or C at positions: 1163, and 1544; nucleotide may be A or T at positions 2058, and 2128; and nucleotide may be A, C, T, or G at position 2251 (see SEQ ID NO: 27).

A primate, e.g., human, SOCS19 nucleotide and corresponding amino acid sequence are provided in SEQ ID NO: 13 and 14. Nucleotide may be A, C, T, or G at positions: 2078, and 2116; and nucleotide may be G or C at position 2063 (see SEQ ID NO: 28).

Finally, a primate, e.g., human, WDS12 nucleotide and corresponding amino acid sequence is provided in SEQ ID NO: 15 and 16. Nucleotide may be A, C, T, or G at positions: 108, and 109; nucleotide may be A or G at positions: 236, 238, and 1258; nucleotide may be G or T at position 233; nucleotide may be G or C at position 234; nucleotide may be C or T at position 237; and nucleotide may be A or T at position 239 (see SEQ ID NO: 29).

SOCS proteins are a family of proteins ranging from approximately 30-60 Kd which inhibit JAK kinase activity. The amino portion of SOCS proteins contain an SH2 binding motif and the carboxy portion of the molecule contains a SOCS box motif which may play a role in dimerization of SOCS proteins. The WDS are closely related in sequence.

SOCS3 expression is induced by IL-2 and can be detected by approximately 1 hour after IL-2 activation. Subsequently, SOCS expression is decreased relatively rapidly (e.g., approximately 8 hrs after activation). Western blots show that SOCS3 interacts with IL-2 receptor and JAK1 following IL-2 stimulation.

II. Definitions

The term "binding composition" refers to molecules that bind with specificity to SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein, e.g., in an antibody-antigen interaction. However, other compounds, e.g., binding proteins, may also specifically associate

with SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 proteins in contrast to other molecules. Typically, the association will be in a natural physiologically relevant protein-protein interaction, either covalent or non-covalent, and may include members of a multiprotein complex, including carrier compounds or dimerization partners. The molecule may be a polymer, or chemical reagent. A functional analog may be a protein with structural modifications, or may be a wholly unrelated molecule, e.g., which has a molecular shape which interacts with the appropriate protein binding determinants. The proteins may serve as agonists or antagonists of the binding partner, see, e.g., Goodman, et al. (eds.) (1990) Goodman & Gilman's: The Pharmacological Bases of Therapeutics (8th ed.) Pergamon Press, Tarrytown, N.Y.

The term "binding agent: SOCS or :WDS protein complex", as used herein, refers to a complex of a binding agent and a SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein that is formed by specific binding of the binding agent to the respective SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein. Specific binding of the binding agent means that the binding agent has a specific binding site that recognizes a site on the SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein. For example, antibodies raised to a SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein and recognizing an epitope on the SOCS or WDS protein are capable of forming a binding agent: SOCS or :WDS protein complex by specific binding. Typically, the formation of a binding agent: SOCS or :WDS protein complex allows the measurement of SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein in a mixture of other proteins and biologics. The term "antibody: SOCS or :WDS protein complex" refers to an embodiment in which the binding agent, e.g., is an antibody. The antibody may be monoclonal, polyclonal, or a binding fragment of an antibody, e.g., an Fv, Fab, or F(ab)2 fragment. The antibody will preferably be a polyclonal antibody for cross-reactivity purposes.

"Homologous" nucleic acid sequences, when compared, exhibit significant similarity, or identity. The standards for homology in nucleic acids are either measures for homology generally used in the art by sequence comparison and/or phylogenetic relationship, or based upon hybridization conditions. Hybridization conditions are described in greater detail below.

An "isolated" nucleic acid is a nucleic acid, e.g., an RNA, DNA, cDNA, genomic DNA, or a mixed polymer, which is substantially separated from other biologic components which naturally accompany a native sequence, e.g., proteins and flanking genomic sequences from the originating species. The term embraces a nucleic acid sequence which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs, or analogs biologically synthesized by heterologous systems. Further, the term includes double-stranded or single-stranded embodiments. Where single-stranded, the nucleic acid may be either the "sense" or the "antisense" strand. A substantially pure molecule includes isolated forms of the molecule. An isolated nucleic acid will usually contain homogeneous nucleic acid molecules, but will, in some embodiments, contain nucleic acids with minor sequence heterogeneity. This heterogeneity is typically found at the polymer ends or portions not critical to a desired biological function or activity.

As used herein, the terms "SOCS" or "WDS" protein shall encompass, when used in a protein context, a protein having amino acid sequences shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, or 16 or a significant fragment of such a protein, preferably a natural embodiment. The term "protein" or "polypeptide" is meant any chain of amino acids, regardless of length or postranslation modification (e.g., glycosylation or phosphorylation). Further, the term encompasses polypeptides which are pre- or pro-proteins. The invention also embraces a polypeptide which exhibits similar structure to SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein, e.g., which interacts with SOCS or WDS protein specific

binding components. These binding components, e.g., antibodies, typically bind to a SOCS or WDS protein, respectively, with high affinity, e.g., at least about 100 nM, usually better than about 30 nM, preferably better than about 10 nM, and more preferably at better than about 3 nM.

The term "polypeptide" or "protein" as used herein includes a significant fragment or segment of a SOCS or WDS protein, and encompasses a stretch of amino acid residues of at least about 8 amino acids, generally at least 10 amino acids, more generally at least 12 amino acids, often at least 14 amino acids, more often at least 16 amino acids, typically at least 18 amino acids, more typically at least 20 amino acids, usually at least 22 amino acids, more usually at least 24 amino acids, preferably at least 26 amino acids, more preferably at least 28 amino acids, and, in particularly preferred embodiments, at least about 30 or more amino acids, e.g., 35, 40, 45, 50, 60, 70, 80, etc. The invention encompasses proteins comprising a plurality of distinct, e.g., nonoverlapping, segments of the specified length. Typically, the plurality will be at least two, more usually at least three, and preferably 5, 7, or even more. While the length minima are provided, longer lengths, of various sizes, may be appropriate, e.g., one of length 7, and two of length 12. Features of one of the different genes should not be taken to limit those of another of the genes.

A "recombinant" nucleic acid is defined either by its method of production or its structure. In reference to its method of production, e.g., a product made by a process, the process is use of recombinant nucleic acid techniques, e.g., involving human intervention in the nucleotide sequence, typically selection or production. Alternatively, it can be a nucleic acid made by generating a sequence comprising fusion of two fragments which are not naturally contiguous to each other, but is meant to exclude products of nature, e.g., naturally occurring mutants. Thus, for example, products made by transforming cells with any non-naturally occurring

vector is encompassed, as are nucleic acids comprising sequence derived using any synthetic oligonucleotide process. Such is often done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a sequence recognition site. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a single genetic entity comprising a desired combination of functions not found in the commonly available natural forms. Restriction enzyme recognition sites are often the target of such artificial manipulations, but other site specific targets, e.g., promoters, DNA replication sites, regulation sequences, control sequences, or other useful features may be incorporated by design. A similar concept is intended for a recombinant, e.g., fusion, polypeptide. Specifically included are synthetic nucleic acids which, by genetic code redundancy, encode polypeptides similar to fragments of these antigens, and fusions of sequences from various different species variants.

"Solubility" is reflected by sedimentation measured in Svedberg units, which are a measure of the sedimentation velocity of a molecule under particular conditions. The determination of the sedimentation velocity was classically performed in an analytical ultracentrifuge, but is typically now performed in a standard ultracentrifuge. See, Freifelder (1982) Physical Biochemistry (2d ed.) W.H. Freeman & Co., San Francisco, CA; and Cantor and Schimmel (1980) Biophysical Chemistry parts 1-3, W.H. Freeman & Co., San Francisco, CA. As a crude determination, a sample containing a putatively soluble polypeptide is spun in a standard full sized ultracentrifuge at about 50K rpm for about 10 minutes, and soluble molecules will remain in the supernatant. A soluble particle or polypeptide will typically be less than about 30S, more typically less than about 15S, usually less than about 10S, more usually less than about 6S, and, in particular embodiments, preferably less than about 4S, and more preferably less than about 3S. Solubility of a polypeptide or fragment

depends upon the environment and the polypeptide. Many parameters affect polypeptide solubility, including temperature, electrolyte environment, size and molecular characteristics of the polypeptide, and nature of the solvent. Typically, the temperature at which the polypeptide is used ranges from about 4° C to about 65° C. Usually the temperature at use is greater than about 18° C and more usually greater than about 22° C. For diagnostic purposes, the temperature will usually be about room temperature or warmer, but less than the denaturation temperature of components in the assay. For therapeutic purposes, the temperature will usually be body temperature, typically about 37° C for humans, though under certain situations the temperature may be raised or lowered in situ or in vitro.

The size and structure of the polypeptide should generally be in a substantially stable state, and usually not in a denatured state. The polypeptide may be associated with other polypeptides in a quaternary structure, e.g., to confer solubility, or associated with lipids or detergents in a manner which approximates natural lipid bilayer interactions.

The solvent will usually be a biologically compatible buffer, of a type used for preservation of biological activities, and will usually approximate a physiological solvent. Usually the solvent will have a neutral pH, typically between about 5 and 10, and preferably about 7.5. On some occasions, a detergent will be added, typically a mild non-denaturing one, e.g., CHS (cholesteryl hemisuccinate) or CHAPS (3-[3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate), or a low enough concentration as to avoid significant disruption of structural or physiological properties of the protein.

"Substantially pure" in a protein context typically means that the protein is isolated from other contaminating proteins, nucleic acids, and other biologicals derived from the original source organism. Purity, or "isolation" may be assayed by standard methods, and will ordinarily be at least about 50% pure,

more ordinarily at least about 60% pure, generally at least about 70% pure, more generally at least about 80% pure, often at least about 85% pure, more often at least about 90% pure, preferably at least about 95% pure, more preferably at least about 98% pure, and in most preferred embodiments, at least 99% pure. Similar concepts apply, e.g., to antibodies or nucleic acids.

"Substantial similarity" in the nucleic acid sequence comparison context means either that the segments, or their complementary strands, when compared, are identical when optimally aligned, with appropriate nucleotide insertions or deletions, in at least about 50% of the nucleotides, generally at least 56%, more generally at least 59%, ordinarily at least 62%, more ordinarily at least 65%, often at least 68%, more often at least 71%, typically at least 74%, more typically at least 77%, usually at least 80%, more usually at least about 85%, preferably at least about 90%, more preferably at least about 95 to 98% or more, and in particular embodiments, as high at about 99% or more of the nucleotides. Alternatively, substantial similarity exists when the segments will hybridize under selective hybridization conditions, to a strand, or its complement, typically using a sequence derived from SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, or 15. Typically, selective hybridization will occur when there is at least about 55% similarity over a stretch of at least about 30 nucleotides, preferably at least about 65% over a stretch of at least about 25 nucleotides, more preferably at least about 75%, and most preferably at least about 90% over about 20 nucleotides. See Kanehisa (1984) Nuc. Acids Res. 12:203-213. The length of similarity comparison, as described, may be over longer stretches, and in certain embodiments will be over a stretch of at least about 17 nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 40 nucleotides, preferably at least about 50 nucleotides, and more preferably at least about 75 to 100 or more nucleotides, e.g., 150, 200, etc.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequent coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optical alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needlman and Wunsch (1970) J. Mol. Biol. 48:443, by the search for similarity method of Pearson and Lipman (1988) Proc. Nat'l Acad. Sci. USA 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (see generally Ausubel et al., supra).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng and Doolittle (1987) J. Mol. Evol. 35:351-360. The method used is similar to the method described by Higgins and Sharp (1989) CABIOS 5:151-153. The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by

a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described Altschul, et al. (1990) J. Mol. Biol. 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http:www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul, et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) Proc. Nat'l Acad. Sci. USA 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993) Proc. Nat'l Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

A further indication that two nucleic acid sequences of polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions, as described below.

"Stringent conditions", in referring to homology or substantial similarity in the hybridization context, will be stringent combined conditions of salt, temperature, organic solvents, and other parameters, typically those controlled in hybridization reactions. The combination of parameters is more important than the measure of any single parameter. See, e.g., Wetmur and Davidson (1968) J. Mol. Biol. 31:349-370.

A nucleic acid probe which binds to a target nucleic acid under stringent conditions is specific for said target nucleic acid. Hybridization under stringent conditions should give a background of at least 2-fold over background, preferably at least 3-5 or more. Such a probe is typically more than 11 nucleotides in length,

and is sufficiently identical or complementary to a target nucleic acid over the region specified by the sequence of the probe to bind the target under stringent hybridization conditions.

SOCS14, SOCS15 (WDS11), SOCS17, SOCS18, SOCS19, or WDS12 protein from other mammalian species can be cloned and isolated by cross-species hybridization of closely related species. See, e.g., below. Similarity may be relatively low between distantly related species, and thus hybridization of relatively closely related species is advisable. Alternatively, preparation of an antibody preparation which exhibits less species specificity may be useful in expression cloning approaches.

The phrase "specifically binds to an antibody" or "specifically immunoreactive with", when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biological components. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not significantly bind other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised to the protein immunogen with the amino acid sequence depicted in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, or 16 can be selected to obtain antibodies specifically immunoreactive with SOCS or WDS proteins and not with other proteins. These antibodies recognize proteins highly similar to the homologous SOCS or WDS protein.

III. Nucleic Acids

Primate or rodent SOCS or WDS protein is each exemplary of a larger class of structurally and functionally related proteins. These soluble proteins will serve to transmit signals between different cell types. The preferred embodiments, as disclosed, will be useful in standard procedures to isolate genes from different individuals or other species, e.g., warm

blooded animals, such as birds and mammals. Cross hybridization will allow isolation of related genes encoding proteins from individuals, strains, or species. A number of different approaches are available to successfully isolate a suitable nucleic acid clone based upon the information provided herein. Southern blot hybridization studies can qualitatively determine the presence of homologous genes in human, monkey, rat, mouse, dog, cow, and rabbit genomes under specific hybridization conditions.

Complementary sequences will also be used as probes or primers. Based upon identification of the likely amino terminus, other peptides should be particularly useful, e.g., coupled with anchored vector or poly-A complementary PCR techniques or with complementary DNA of other peptides.

Techniques for nucleic acid manipulation of genes encoding SOCS or WDS proteins, such as subcloning nucleic acid sequences encoding polypeptides into expression vectors, labeling probes, DNA hybridization, and the like are described generally in Sambrook, et al. (1989) Molecular Cloning: A Laboratory Manual (2nd ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY, which is incorporated herein by reference. This manual is hereinafter referred to as "Sambrook, et al."

There are various methods of isolating DNA sequences encoding SOCS or WDS proteins. For example, DNA is isolated from a genomic or cDNA library using labeled oligonucleotide probes having sequences identical or complementary to the sequences disclosed herein. Full-length probes may be used, or oligonucleotide probes may be generated by comparison of the sequences disclosed. Such probes can be used directly in hybridization assays to isolate DNA encoding SOCS or WDS proteins, or probes can be designed for use in amplification techniques such as PCR, for the isolation of DNA encoding SOCS or WDS proteins.

To prepare a cDNA library, mRNA is isolated from cells which expresses a SOCS or WDS protein. cDNA is

prepared from the mRNA and ligated into a recombinant vector. The vector is transfected into a recombinant host for propagation, screening, and cloning. Methods for making and screening cDNA libraries are well known. See Gubler and Hoffman (1983) Gene 25:263-269 and Sambrook, et al.

For a genomic library, the DNA can be extracted from tissue and either mechanically sheared or enzymatically digested to yield fragments of about 12-20 kb. The fragments are then separated by gradient centrifugation and cloned in bacteriophage lambda vectors. These vectors and phage are packaged in vitro, as described in Sambrook, et al. Recombinant phage are analyzed by plaque hybridization as described in Benton and Davis (1977) Science 196:180-182. Colony hybridization is carried out as generally described in e.g., Grunstein, et al. (1975) Proc. Natl. Acad. Sci. USA. 72:3961-3965.

DNA encoding a SOCS14 or SOCS15 protein can be identified in either cDNA or genomic libraries by its ability to hybridize with the nucleic acid probes described herein, e.g., in colony or plaque hybridization assays. The corresponding DNA regions are isolated by standard methods familiar to those of skill in the art. See, e.g., Sambrook, et al.

Various methods of amplifying target sequences, such as the polymerase chain reaction, can also be used to prepare DNA encoding SOCS or WDS proteins. Polymerase chain reaction (PCR) technology is used to amplify such nucleic acid sequences directly from mRNA, from cDNA, and from genomic libraries or cDNA libraries. The isolated sequences encoding SOCS or WDS proteins may also be used as templates for PCR amplification.

Typically, in PCR techniques, oligonucleotide primers complementary to two 5' regions in the DNA region to be amplified are synthesized. The polymerase chain reaction is then carried out using the two primers. See Innis, et al. (eds.) (1990) PCR Protocols: A Guide to Methods and Applications Academic Press, San Diego, CA. Primers can be selected to amplify the entire regions encoding a full-length SOCS or WDS protein or to amplify

smaller DNA segments as desired. Once such regions are PCR-amplified, they can be sequenced and oligonucleotide probes can be prepared from sequence obtained using standard techniques. These probes can then be used to isolate DNA's encoding SOCS or WDS proteins.

Oligonucleotides for use as probes are usually chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage and Carruthers (1983) Tetrahedron Lett. 22(20):1859-1862, or using an automated synthesizer, as described in Needham-VanDevanter, et al. (1984) Nucleic Acids Res. 12:6159-6168. Purification of oligonucleotides is performed e.g., by native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson and Regnier (1983) J. Chrom. 255:137-149. The sequence of the synthetic oligonucleotide can be verified using, e.g., the chemical degradation method of Maxam, A.M. and Gilbert, W. in Grossman, L. and Moldave (eds.) (1980) Methods in Enzymology 65:499-560 Academic Press, New York.

Isolated nucleic acids encoding SOCS or WDS proteins were identified.The nucleotide sequences and corresponding open reading frames are provided in SEQ ID NO: 1 through 16.

These SOCS or WDS proteins exhibit limited similarity to portions other intracellular proteins. In particular, β-sheet and α-helix residues can be determined using, e.g., RASMOL program, see Sayle and Milner-White (1995) TIBS 20:374-376; or Gronenberg, et al. (1991) Protein Engineering 4:263-269; and other structural features are defined in Lodi, et al. (1994) Science 263:1762-1767.

This invention provides isolated DNA or fragments to encode a SOCS or WDS protein. In addition, this invention provides isolated or recombinant DNA which encodes a protein or polypeptide which is capable of hybridizing under appropriate conditions, e.g., high stringency, with the DNA sequences described herein. Said biologically active protein or polypeptide can be an intact protein, or fragment, and have an amino acid

sequence as disclosed in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, or 16, particularly natural embodiments. Preferred embodiments will be full length natural sequences. Further, this invention contemplates the use of isolated or recombinant DNA, or fragments thereof, which encode proteins which are homologous to a SOCS or WDS protein or which were isolated using cDNA encoding a SOCS or WDS protein as a probe. The isolated DNA can have the respective regulatory sequences in the 5' and 3' flanks, e.g., promoters, enhancers, poly-A addition signals, and others. Also embraced are methods for making expression vectors with these sequences, or for making, e.g., expressing and purifying, protein products.

A DNA which codes for a SOCS or WDS protein will be particularly useful to identify genes, mRNA, and cDNA species which code for related or similar proteins, as well as DNAs which code for homologous proteins from different species. There are likely homologs in other species, including primates, rodents, canines, felines, and birds. Various SOCS or WDS proteins should be homologous and are encompassed herein. However, even proteins that have a more distant evolutionary relationship to the antigen can readily be isolated under appropriate conditions using these sequences if they are sufficiently homologous. Primate SOCS or WDS proteins are of particular interest.

Recombinant clones derived from the genomic sequences, e.g., containing introns, will be useful for transgenic studies, including, e.g., transgenic cells and organisms, and for gene therapy. See, e.g., Goodnow (1992) "Transgenic Animals" in Roitt (ed.) Encyclopedia of Immunology, Academic Press, San Diego, pp. 1502-1504; Travis (1992) Science 256:1392-1394; Kuhn, et al. (1991) Science 254:707-710; Capecchi (1989) Science 244:1288; Robertson (1987)(ed.) Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, IRL Press, Oxford; and Rosenberg (1992) J. Clinical Oncology 10:180-199.

IV. Antibodies

Antibodies can be raised to various SOCS14 or SOCS15 proteins, including individual, polymorphic, allelic, strain, or species variants, and fragments thereof, both in their naturally occurring (full-length) forms and in their recombinant forms. Additionally, antibodies can be raised to SOCS or WDS proteins in either their active forms or in their inactive forms. Anti-idiotypic antibodies may also be used.

A. Antibody Production

A number of immunogens may be used to produce antibodies specifically reactive with SOCS or WDS proteins. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring protein may also be used either in pure or impure form. Synthetic peptides, made using the human SOCS14 or SOCS15 protein sequences described herein, may also used as an immunogen for the production of antibodies to SOCS14 or SOCS15 proteins. Recombinant protein can be expressed in eukaryotic or prokaryotic cells as described herein, and purified as described. Naturally folded or denatured material can be used, as appropriate, for producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the protein.

Methods of producing polyclonal antibodies are known to those of skill in the art. Typically, an immunogen, preferably a purified protein, is mixed with an adjuvant and animals are immunized with the mixture. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the SOCS or WDS protein of interest. When appropriately high titers of antibody to the immunogen are obtained, usually after repeated immunizations, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired. See, e.g., Harlow and Lane; or Coligan.

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art.

Typically, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (see, Kohler and Milstein (1976) Eur. J. Immunol. 6:511-519, incorporated herein by reference). Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according, e.g., to the general protocol outlined by Huse, et al. (1989) Science 246:1275-1281.

Antibodies, including binding fragments and single chain versions, against predetermined fragments of SOCS or WDS protein can be raised by immunization of animals with conjugates of the fragments with carrier proteins as described above. Monoclonal antibodies are prepared from cells secreting the desired antibody. These antibodies can be screened for binding to normal or defective SOCS or WDS proteins, or screened for agonistic or antagonistic activity, e.g., effect on cell cycle progression or transcription of specific genes. These monoclonal antibodies will usually bind with at least a K_D of about 1 mM, more usually at least about 300 µM, typically at least about 10 µM, more typically at least about 30 µM, preferably at least about 10 µM, and more preferably at least about 3 µM or better.

In some instances, it is desirable to prepare monoclonal antibodies from various mammalian hosts, such as mice, rodents, primates, humans, etc. Description of techniques for preparing such monoclonal antibodies may be found in, e.g., Stites, et al. (eds.) Basic and Clinical Immunology (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and

Lane (1988) Antibodies: A Laboratory Manual CSH Press; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.) Academic Press, New York, NY; and particularly in Kohler and Milstein (1975) Nature 256:495-497, which discusses one method of generating monoclonal antibodies. Summarized briefly, this method involves injecting an animal with an immunogen. The animal is then sacrificed and cells taken from its spleen, which are then fused with myeloma cells. The result is a hybrid cell or "hybridoma" that is capable of reproducing in vitro. The population of hybridomas is then screened to isolate individual clones, each of which secrete a single antibody species to the immunogen. In this manner, the individual antibody species obtained are the products of immortalized and cloned single B cells from the immune animal generated in response to a specific site recognized on the immunogenic substance.

Other suitable techniques involve selection of libraries of antibodies in phage or similar vectors. See, e.g., Huse, et al. (1989) "Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire in Phage Lambda," Science 246:1275-1281; and Ward, et al. (1989) Nature 341:544-546. The polypeptides and antibodies of the present invention may be used with or without modification, including chimeric or humanized antibodies. Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents, teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins may be produced. See, Cabilly, U.S.

Patent No. 4,816,567; and Queen, et al. (1989) Proc. Nat'l Acad. Sci. USA 86:10029-10033.

The antibodies of this invention are useful for affinity chromatography in isolating SOCS or WDS protein. Columns can be prepared where the antibodies are linked to a solid support, e.g., particles, such as agarose, SEPHADEX, or the like, where a cell lysate or supernatant may be passed through the column, the column washed, followed by increasing concentrations of a mild denaturant, whereby purified SOCS or WDS protein will be released.

The antibodies may also be used to screen expression libraries for particular expression products. Usually the antibodies used in such a procedure will be labeled with a moiety allowing easy detection of presence of antigen by antibody binding.

Antibodies to SOCS or WDS proteins may be used for the identification of cell populations expressing the proteins. By assaying, e.g., by histology or otherwise, probably a disruptive assay which kills that sample of cells, the expression products of cells expressing SOCS or WDS proteins it is possible to diagnose disease, e.g., cancerous conditions.

Antibodies raised against each SOCS or WDS protein will also be useful to raise anti-idiotypic antibodies. These will be useful in detecting or diagnosing various immunological conditions related to expression of the respective antigens.

B. Immunoassays

A particular protein can be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures in general, see Stites and Terr (eds.) (1991) Basic and Clinical Immunology (7th ed.). Moreover, the immunoassays of the present invention can be performed in many configurations, which are reviewed extensively in Maggio (ed.) (1980) Enzyme Immunoassay CRC Press, Boca Raton, Florida; Tijan (1985) "Practice and Theory of Enzyme Immunoassays," Laboratory Techniques in

Biochemistry and Molecular Biology, Elsevier Science Publishers B.V., Amsterdam; and Harlow and Lane Antibodies, A Laboratory Manual, supra, each of which is incorporated herein by reference. See also Chan (ed.) (1987) Immunoassay: A Practical Guide Academic Press, Orlando, FL; Price and Newman (eds.) (1991) Principles and Practice of Immunoassays Stockton Press, NY; and Ngo (ed.) (1988) Non-isotopic Immunoassays Plenum Press, NY.

Immunoassays for measurement of SOCS or WDS proteins can be performed by a variety of methods known to those skilled in the art. In brief, immunoassays to measure the protein can be either competitive or noncompetitive binding assays. In competitive binding assays, the sample to be analyzed competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is an antibody specifically reactive with SOCS or WDS proteins produced as described above. The concentration of labeled analyte bound to the capture agent is inversely proportional to the amount of free analyte present in the sample.

In a competitive binding immunoassay, the SOCS or WDS protein present in the sample competes with labeled protein for binding to a specific binding agent, for example, an antibody specifically reactive with the SOCS or WDS protein. The binding agent may be bound to a solid surface to effect separation of bound labeled protein from the unbound labeled protein. Alternately, the competitive binding assay may be conducted in liquid phase and a variety of techniques known in the art may be used to separate the bound labeled protein from the unbound labeled protein. Following separation, the amount of bound labeled protein is determined. The amount of protein present in the sample is inversely proportional to the amount of labeled protein binding.

Alternatively, a homogeneous immunoassay may be performed in which a separation step is not needed. In these immunoassays, the label on the protein is altered by the binding of the protein to its specific binding

agent. This alteration in the labeled protein results in a decrease or increase in the signal emitted by label, so that measurement of the label at the end of the immunoassay allows for detection or quantitation of the protein.

Qualitative or quantitative analysis of SOCS or WDS proteins may also be determined by a variety of noncompetitive immunoassay methods. For example, a two-site, solid phase sandwich immunoassay may be used. In this type of assay, a binding agent for the protein, for example an antibody, is attached to a solid support. A second protein binding agent, which may also be an antibody, and which binds the protein at a different site, is labeled. After binding at both sites on the protein has occurred, the unbound labeled binding agent is removed and the amount of labeled binding agent bound to the solid phase is measured. The amount of labeled binding agent bound is directly proportional to the amount of protein in the sample.

Western blot analysis can be used to determine the presence of SOCS or WDS proteins in a sample. Electrophoresis is carried out, for example, on a tissue sample suspected of containing the protein. Following electrophoresis to separate the proteins, and transfer of the proteins to a suitable solid support, e.g., a nitrocellulose filter, the solid support is incubated with an antibody reactive with the protein. This antibody may be labeled, or alternatively may be detected by subsequent incubation with a second labeled antibody that binds the primary antibody.

The immunoassay formats described above employ labeled assay components. The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. A wide variety of labels and methods may be used. Traditionally, a radioactive label incorporating ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P was used. Non-radioactive labels include proteins which bind to labeled antibodies, fluorophores, chemiluminescent agents, enzymes, and antibodies which can serve as specific binding pair

members for a labeled protein. The choice of label depends on sensitivity required, ease of conjugation with the compound, stability requirements, and available instrumentation. For a review of various labeling or signal producing systems which may be used, see U.S. Patent No. 4,391,904, which is incorporated herein by reference.

Antibodies reactive with a particular protein can also be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures applicable to the measurement of antibodies by immunoassay techniques, see Stites and Terr (eds.) Basic and Clinical Immunology (7th ed.) supra; Maggio (ed.) Enzyme Immunoassay, supra; and Harlow and Lane Antibodies, A Laboratory Manual, supra.

In brief, immunoassays to measure antisera reactive with SOCS or WDS proteins can be either competitive or noncompetitive binding assays. In competitive binding assays, the sample analyte competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is a purified recombinant SOCS or WDS protein produced as described above. Other sources of these proteins, including isolated or partially purified naturally occurring protein, may also be used. Noncompetitive assays include sandwich assays, in which the sample analyte is bound between two analyte-specific binding reagents. One of the binding agents is used as a capture agent and is bound to a solid surface. The second binding agent is labeled and is used to measure or detect the resultant complex by visual or instrument means. A number of combinations of capture agent and labeled binding agent can be used. A variety of different immunoassay formats, separation techniques, and labels can be also be used similar to those described above for the measurement of SOCS or WDS proteins.

V. Making SOCS or WDS proteins; Mimetics

DNAs which encode a SOCS or WDS protein or fragments thereof can be obtained by chemical synthesis, screening

cDNA libraries, or by screening genomic libraries prepared from a wide variety of cell lines or tissue samples. Methods for doing so, or making expression vectors are described herein.

These DNAs can be expressed in a wide variety of host cells for the synthesis of a full-length protein or fragments which can in turn, e.g., be used to generate polyclonal or monoclonal antibodies; for binding studies; for construction and expression of modified molecules; and for structure/function studies. Each SOCS or WDS protein or its fragments can be expressed in host cells that are transformed or transfected with appropriate expression vectors. By "transformed" is meant a cell into which (or into an ancestor of which) has been introduced, by means of recombinant techniques, a DNA molecule that encodes a SOCS or WDS polypeptide. Heterologously expressed SOCS or WDS polypeptides can be substantially purified to be free of protein or cellular contaminants, other than those derived from the recombinant host, and therefore are particularly useful in pharmaceutical compositions when combined with a pharmaceutically acceptable carrier and/or diluent. The antigen, e.g., SOCS or WDS protein, or portions thereof, may be expressed as fusions with other proteins or possessing an epitope tag.

Expression vectors are typically self-replicating DNA or RNA constructs containing the desired antigen gene or its fragments, usually operably linked to appropriate genetic control elements that are recognized in a suitable host cell. The specific type of control elements necessary to effect expression will depend upon the eventual host cell used. Generally, the genetic control elements can include a prokaryotic promoter system or a eukaryotic promoter expression control system, and typically include a transcriptional promoter, an optional operator to control the onset of transcription, transcription enhancers to elevate the level of mRNA expression, a sequence that encodes a suitable ribosome binding site, and sequences that terminate transcription and translation. All of the

associated elements both necessary and sufficient for the production of SOCS or WDS polypeptide will be in operable linkage with the nucleic acid encoding a SOCS or WDS polypeptide. Expression vectors also usually contain an origin of replication that allows the vector to replicate independently from the host cell.

The vectors of this invention contain DNAs which encode a SOCS or WDS protein, or a fragment thereof, typically encoding, e.g., a biologically active polypeptide, or protein. The DNA can be under the control of a viral promoter and can encode a selection marker. This invention further contemplates use of such expression vectors which are capable of expressing eukaryotic cDNA coding for a SOCS or WDS protein in a prokaryotic or eukaryotic host, where the vector is compatible with the host and where the eukaryotic cDNA coding for the protein is inserted into the vector such that growth of the host containing the vector expresses the cDNA in question. Usually, expression vectors are designed for stable replication in their host cells or for amplification to greatly increase the total number of copies of the desirable gene per cell. It is not always necessary to require that an expression vector replicate in a host cell, e.g., it is possible to effect transient expression of the protein or its fragments in various hosts using vectors that do not contain a replication origin that is recognized by the host cell. It is also possible to use vectors that cause integration of a SOCS or WDS protein gene or its fragments into the host DNA by recombination, or to integrate a promoter which controls expression of an endogenous gene.

Vectors, as used herein, contemplate plasmids, viruses, bacteriophage, integratable DNA fragments, and other vehicles which enable the integration of DNA fragments into the genome of the host. Expression vectors are specialized vectors which contain genetic control elements that effect expression of operably linked genes. Plasmids are the most commonly used form of vector, but many other forms of vectors which serve an equivalent function are suitable for use herein. See,

e.g., Pouwels, et al. (1985 and Supplements) Cloning Vectors: A Laboratory Manual Elsevier, N.Y.; and Rodriquez, et al. (eds.) (1988) Vectors: A Survey of Molecular Cloning Vectors and Their Uses Buttersworth, Boston, MA.

Suitable host cells include prokaryotes, lower eukaryotes, and higher eukaryotes. Prokaryotes include both gram negative and gram positive organisms, e.g., E. coli and B. subtilis. Lower eukaryotes include yeasts, e.g., S. cerevisiae and Pichia, and species of the genus Dictyostelium. Higher eukaryotes include established tissue culture cell lines from animal cells, both of non-mammalian origin, e.g., insect cells, and birds, and of mammalian origin, e.g., human, primates, and rodents.

Prokaryotic host-vector systems include a wide variety of vectors for many different species. As used herein, E. coli and its vectors will be used generically to include equivalent vectors used in other prokaryotes. A representative vector for amplifying DNA is pBR322 or its derivatives. Vectors that can be used to express these proteins or protein fragments include, but are not limited to, such vectors as those containing the lac promoter (pUC-series); trp promoter (pBR322-trp); Ipp promoter (the pIN-series); lambda-pP or pR promoters (pOTS); or hybrid promoters such as ptac (pDR540). See Brosius, et al. (1988) "Expression Vectors Employing Lambda-, trp-, lac-, and Ipp-derived Promoters", in Rodriguez and Denhardt (eds.) Vectors: A Survey of Molecular Cloning Vectors and Their Uses 10:205-236 Buttersworth, Boston, MA.

Lower eukaryotes, e.g., yeasts and Dictyostelium, may be transformed with SOCS or WDS protein sequence containing vectors. For purposes of this invention, the most common lower eukaryotic host is the baker's yeast, Saccharomyces cerevisiae. It will be used generically to represent lower eukaryotes although a number of other strains and species are also available. Yeast vectors typically consist of a replication origin (unless of the integrating type), a selection gene, a promoter, DNA

encoding the desired protein or its fragments, and sequences for translation termination, polyadenylation, and transcription termination. Suitable expression vectors for yeast include such constitutive promoters as 3-phosphoglycerate kinase and various other glycolytic enzyme gene promoters or such inducible promoters as the alcohol dehydrogenase 2 promoter or metallothionine promoter. Suitable vectors include derivatives of the following types: self-replicating low copy number (such as the YRp-series), self-replicating high copy number (such as the YEp-series); integrating types (such as the YIp-series), or mini-chromosomes (such as the YCp-series).

Higher eukaryotic tissue culture cells are typically the preferred host cells for expression of the functionally active SOCS or WDS protein. In principle, many higher eukaryotic tissue culture cell lines may be used, e.g., insect baculovirus expression systems, whether from an invertebrate or vertebrate source. However, mammalian cells are preferred to achieve proper processing, both cotranslationally and posttranslationally. Transformation or transfection and propagation of such cells is routine. Useful cell lines include HeLa cells, Chinese hamster ovary (CHO) cell lines, baby rat kidney (BRK) cell lines, insect cell lines, bird cell lines, and monkey (COS) cell lines. Expression vectors for such cell lines usually include an origin of replication, a promoter, a translation initiation site, RNA splice sites (e.g., if genomic DNA is used), a polyadenylation site, and a transcription termination site. These vectors also may contain a selection gene or amplification gene. Suitable expression vectors may be plasmids, viruses, or retroviruses carrying promoters derived, e.g., from such sources as from adenovirus, SV40, parvoviruses, vaccinia virus, or cytomegalovirus. Representative examples of suitable expression vectors include pCDNA1; pCD, see Okayama, et al. (1985) Mol. Cell Biol. 5:1136-1142; pMC1neo Poly-A, see Thomas, et al. (1987) Cell 51:503-512; and a baculovirus vector such as pAC 373 or pAC 610.

It is likely that SOCS or WDS proteins need not be glycosylated to elicit biological responses. However, it will occasionally be desirable to express a SOCS or WDS protein polypeptide in a system which provides a specific or defined glycosylation pattern. In this case, the usual pattern will be that provided naturally by the expression system. However, the pattern will be modifiable by exposing the polypeptide, e.g., in unglycosylated form, to appropriate glycosylating proteins introduced into a heterologous expression system. For example, the SOCS or WDS protein gene may be co-transformed with one or more genes encoding mammalian or other glycosylating enzymes. It is further understood that over glycosylation may be detrimental to SOCS or WDS protein biological activity, and that one of skill may perform routine testing to optimize the degree of glycosylation which confers optimal biological activity.

Furthermore, heterologously expressed proteins or polypeptides can also be expressed in plant cells. For plant cells viral expression vectors (e.g., cauliflower mosaic virus and tobacco mosaic virus) and plasmid expression vectors (e.g., T1 plasmid) are suitable. Such cells are available from a wide range of sources (e.g., the American Tissue Type Culture Collection, Rockland, MD; also, see for example, Ausubel, et al. (cur. ed. and Supplements; expression vehicles may be chosen from those provided e.g., in Pouwels, et al. (Cur. ed..) Cloning Vectors, A Laboratory Manual).

A SOCS or WDS protein, or a fragment thereof, may be engineered to be phosphatidyl inositol (PI) linked to a cell membrane, but can be removed from membranes by treatment with a phosphatidyl inositol cleaving enzyme, e.g., phosphatidyl inositol phospholipase-C. This releases the antigen in a biologically active form, and allows purification by standard procedures of protein chemistry. See, e.g., Low (1989) Biochem. Biophys. Acta 988:427-454; Tse, et al. (1985) Science 230:1003-1008; and Brunner, et al. (1991) J. Cell Biol. 114:1275-1283.

Now that SOCS or WDS proteins have been characterized, fragments or derivatives thereof can be

prepared by conventional processes for synthesizing peptides. These include processes such as are described in Stewart and Young (1984) Solid Phase Peptide Synthesis Pierce Chemical Co., Rockford, IL; Bodanszky and Bodanszky (1984) The Practice of Peptide Synthesis Springer-Verlag, New York, NY; and Bodanszky (1984) The Principles of Peptide Synthesis Springer-Verlag, New York, NY. For example, an azide process, an acid chloride process, an acid anhydride process, a mixed anhydride process, an active ester process (for example, p-nitrophenyl ester, N-hydroxysuccinimide ester, or cyanomethyl ester), a carbodiimidazole process, an oxidative-reductive process, or a dicyclohexylcarbodiimide (DCCD)/additive process can be used. Solid phase and solution phase syntheses are both applicable to the foregoing processes.

The prepared protein and fragments thereof can be isolated and purified from the reaction mixture by means of peptide separation, for example, by extraction, precipitation, electrophoresis and various forms of chromatography, and the like. The SOCS or WDS proteins of this invention can be obtained in varying degrees of purity depending upon its desired use. Purification can be accomplished by use of known protein purification techniques or by the use of the antibodies or binding partners herein described, e.g., in immunoabsorbant affinity chromatography. This immunoabsorbant affinity chromatography is carried out by first linking the antibodies to a solid support and then contacting the linked antibodies with solubilized lysates of appropriate source cells, lysates of other cells expressing the protein, or lysates or supernatants of cells producing the SOCS or WDS proteins as a result of recombinant DNA techniques, see below.

Multiple cell lines may be screened for one which expresses a SOCS or WDS protein at a high level compared with other cells. Various cell lines, e.g., a mouse thymic stromal cell line TA4, is screened and selected for its favorable handling properties. Natural SOCS or WDS proteins can be isolated from natural sources, or by

expression from a transformed cell using an appropriate expression vector. Purification of the expressed protein is achieved by standard procedures, or may be combined with engineered means for effective purification at high efficiency from cell lysates or supernatants. Epitope or other tags, e.g., FLAG or His_6 segments, can be used for such purification features.

VI. Physical Variants

This invention also encompasses proteins or peptides having substantial amino acid sequence similarity with an amino acid sequence of a SOCS or WDS protein. Natural variants include individual, polymorphic, allelic, strain, or species variants.

Amino acid sequence similarity, or sequence identity, is determined by optimizing residue matches, if necessary, by introducing gaps as required. This changes when considering conservative substitutions as matches. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Homologous amino acid sequences include natural polymorphic, allelic, and interspecies variations in each respective protein sequence. Typical homologous proteins or peptides will have from 50-100% similarity (if gaps can be introduced), to 75-100% similarity (if conservative substitutions are included) over fixed stretches of amino acids with the amino acid sequence of the SOCS or WDS protein. Similarity measures will be at least about 50%, generally at least 65%, usually at least 70%, preferably at least 75%, and more preferably at least 90%, and in particularly preferred embodiments, at least 96% or more. See also Needleham, et al. (1970) J. Mol. Biol. 48:443-453; Sankoff, et al. (1983) Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison Chapter One, Addison-Wesley, Reading, MA; and software packages from IntelliGenetics, Mountain View, CA; and the University of Wisconsin

Genetics Computer Group, Madison, WI. Stretches of amino acids will be at least about 10 amino acids, usually about 20 amino acids, usually 50 amino acids, preferably 75 amino acids, and in particularly preferred embodiments at least about 100 amino acids. Identity can also be measures over amino acid stretches of about 98, 99, 110, 120, 130, etc.

Nucleic acids encoding mammalian SOCS or WDS proteins will typically hybridize to the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, or 15 under stringent conditions. For example, nucleic acids encoding human SOCS or WDS proteins will normally hybridize to the nucleic acid of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, or 15 under stringent hybridization conditions. Generally, stringent conditions are selected to be about 10° C lower than the thermal melting point (Tm) for the probe sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is about 0.2 molar at pH 7 and the temperature is at least about 50° C. Other factors may significantly affect the stringency of hybridization, including, among others, base composition and size of the complementary strands, the presence of organic solvents such as formamide, and the extent of base mismatching. A preferred embodiment will include nucleic acids which will bind to disclosed sequences in 50% formamide and 200 mM NaCl at 42° C.

Hybridizing nucleic acids to SOCS nucleic acid of the invention can be used as a cloning probe, a primer (e.g., a PCR primer), or a diagnostic probe. Hybridizing nucleic acids can be splice variants encoded by one of the SOCS genes described herein. Thus, the hybridizing nucleic acids may encode a polypeptide that is shorter or longer than the various forms of SOCS described herein. Hybridizing nucleic acids may also encode proteins that are related to SOCS (e.g., polypeptides encoded by genes

that include a portion having a relatively high degree of identity to a SOCS gene described herein).

An isolated SOCS or WDS protein encoding DNA can be readily modified by nucleotide substitutions, nucleotide deletions, nucleotide insertions, and short inversions of nucleotide stretches. These modifications result in novel DNA sequences which encode SOCS or WDS protein antigens, their derivatives, or proteins having highly similar physiological, immunogenic, or antigenic activity.

Modified sequences can be used to produce mutant antigens or to enhance expression. Enhanced expression may involve gene amplification, increased transcription, increased translation, and other mechanisms. Such mutant SOCS or WDS protein derivatives include predetermined or site-specific mutations of the respective protein or its fragments. "Mutant SOCS or WDS protein" encompasses a polypeptide otherwise falling within the homology definition of the human or rodent SOCS or WDS protein as set forth above, but having an amino acid sequence which differs from that of a SOCS or WDS protein as found in nature, whether by way of deletion, substitution, or insertion. In particular, "site specific mutant SOCS or WDS protein" generally includes proteins having significant similarity with a protein having a sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, or 16, e.g., natural embodiments, and as sharing various biological activities, e.g., antigenic or immunogenic, with those sequences, and in preferred embodiments contain most or all of the disclosed sequence. This applies also to polymorphic variants from different individuals. Similar concepts apply to different SOCS or WDS proteins, particularly those found in various warm blooded animals, e.g., mammals and birds. As stated before, it is emphasized that descriptions are generally meant to encompass other SOCS or WDS proteins, not limited to the human embodiments specifically discussed.

The invention encompasses, but is not limited to, SOCS proteins and polypeptides that are functionally related to SOCS encoded by the specific sequence

identifiers of the present application. Functionally related proteins and polypeptides include any protein or polypeptide sharing a functional characteristic with SOCS of the present invention e.g., the ability to interact with Janus family tyrosine kinases or the ability to be induced by IL-2 receptor activation. Such functionally related SOCS polypeptides include, but are not limited to, additions or substitutions of amino acid residues within the amino acid sequence encoded by the SOCS sequences described herein which result in a silent change, thus producing a functionally equivalent SOCS polypeptide. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphiphatic nature of the residues involved.

For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

While random mutations can be made to SOCS nucleic acid (using well known random mutagenesis techniques) and the resulting SOCS polypeptides can be tested for activity, site-directed mutations of SOCS coding sequences can be engineered (using well known site-directed mutagenesis techniques) to generate mutant SOCS with increased function, e.g. greater inhibition of JANUS kinase activity or greater resistance to degradation.

To design functionally related and functionally variant SOCS polypeptides, it is useful to distinguish between conserved and variable amino residues using the homology comparison tables provided herein.

To preserve SOCS function, it is preferable that conserved residues remain unaltered and that the conformational folding of the SOCS functional sites be preserved. Preferably, alteration of non-conserved residues are carried out with conservative alterations

e.g., a basic amino acid is replaced by a different basic amino acid. To produce altered function variants, it is preferred to make non-conservative changes at variable and or conserved residues. Deletions at conserved and variable residues can also be used to create altered function variants.

Although site specific mutation sites are predetermined, mutants need not be site specific. SOCS or WDS protein mutagenesis can be conducted by making amino acid insertions or deletions. Substitutions, deletions, insertions, or any combinations may be generated to arrive at a final construct. Insertions include amino- or carboxyl- terminal fusions, e.g. epitope tags. Random mutagenesis can be conducted at a target codon and the expressed mutants can then be screened for the desired activity. Methods for making substitution mutations at predetermined sites in DNA having a known sequence are well known in the art, e.g., by M13 primer mutagenesis or polymerase chain reaction (PCR) techniques. See also, Sambrook, et al. (1989) and Ausubel, et al. (1987 and Supplements). The mutations in the DNA normally should not place coding sequences out of reading frames and preferably will not create complementary regions that could hybridize to produce secondary mRNA structure such as loops or hairpins.

The present invention also provides recombinant proteins, e.g., heterologous fusion proteins using segments from these proteins. A heterologous fusion protein is a fusion of proteins or segments which are naturally not normally fused in the same manner e.g., a marker polypeptide or fusion partner. For example, the polypeptide can be fused to a hexa-histidine tag to facilitate purification or bacterially expressed protein or a hemaglutinin tag to facilitate purification or protein expressed in eukaryotic cells. Thus, the fusion product of an immunoglobulin with a SOCS or WDS protein polypeptide is a continuous protein molecule having sequences fused in a typical peptide linkage, typically made as a single translation product and exhibiting

properties derived from each source peptide. A similar concept applies to heterologous nucleic acid sequences.

In addition, new constructs may be made from combining similar functional domains from other proteins. For example, protein-binding or other segments may be "swapped" between different new fusion polypeptides or fragments. See, e.g., Cunningham, et al. (1989) Science 243:1330-1336; and O'Dowd, et al. (1988) J. Biol. Chem. 263:15985-15992. Thus, new chimeric polypeptides exhibiting new combinations of specificities will result from the functional linkage of protein-binding specificities and other functional domains.

VII. Functional Variants

The blocking of physiological response to SOCS or WDS protein may result from the inhibition of binding of the protein to its binding partner, e.g., through competitive inhibition. Thus, in vitro assays of the present invention will often use isolated protein, membranes from cells expressing a recombinant membrane associated SOCS or WDS protein, soluble fragments comprising binding segments of these proteins, or fragments attached to solid phase substrates. These assays will also allow for the diagnostic determination of the effects of either binding segment mutations and modifications, or protein mutations and modifications, e.g., protein analogs. This invention also contemplates the use of competitive drug screening assays, e.g., where neutralizing antibodies to antigen or binding partner fragments compete with a test compound for binding to the protein. In this manner, the antibodies can be used to detect the presence of a polypeptide which shares one or more antigenic binding sites of the protein and can also be used to occupy binding sites on the protein that might otherwise interact with a binding partner.

"Derivatives" of SOCS or WDS protein antigens include amino acid sequence mutants, glycosylation variants, and covalent or aggregate conjugates with other chemical moieties. Covalent derivatives can be prepared by linkage of functionalities to groups which are found

in SOCS or WDS protein amino acid side chains or at the N- or C- termini, by means which are well known in the art. These derivatives can include, without limitation, aliphatic esters or amides of the carboxyl terminus, or of residues containing carboxyl side chains, O-acyl derivatives of hydroxyl group-containing residues, and N-acyl derivatives of the amino terminal amino acid or amino-group containing residues, e.g., lysine or arginine. Acyl groups are selected from the group of alkyl-moieties including C3 to C18 normal alkyl, thereby forming alkanoyl aroyl species. Covalent attachment to carrier proteins may be important when immunogenic moieties are haptens.

In particular, glycosylation alterations are included, e.g., made by modifying the glycosylation patterns of a polypeptide during its synthesis and processing, or in further processing steps. Particularly preferred means for accomplishing this are by exposing the polypeptide to glycosylating enzymes derived from cells which normally provide such processing, e.g., mammalian glycosylation enzymes. Deglycosylation enzymes are also contemplated. Also embraced are versions of the same primary amino acid sequence which have other minor modifications, including phosphorylated amino acid residues, e.g., phosphotyrosine, phosphoserine, or phosphothreonine, or other moieties, including ribosyl groups or cross-linking reagents.

A major group of derivatives are covalent conjugates of the SOCS or WDS protein or fragments thereof with other proteins or polypeptides. These derivatives can be synthesized in recombinant culture such as N- or C-terminal fusions or by the use of agents known in the art for their usefulness in cross-linking proteins through reactive side groups. Preferred protein derivatization sites with cross-linking agents are at free amino groups, carbohydrate moieties, and cysteine residues.

Fusion polypeptides between SOCS or WDS protein and other homologous or heterologous proteins are also provided. Heterologous polypeptides may be fusions between different surface markers, resulting in, e.g., a

hybrid protein exhibiting binding partner specificity. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative proteins. Typical examples are fusions of a reporter polypeptide, e.g., luciferase, with a segment or domain of a protein, e.g., a segment involved in binding partner interaction, so that the presence or location of the fused protein may be easily determined. See, e.g., Dull, et al., U.S. Patent No. 4,859,609. Other gene fusion partners include bacterial β-galactosidase, trpE, Protein A, ß-lactamase, alpha amylase, alcohol dehydrogenase, and yeast alpha mating factor. See, e.g., Godowski, et al. (1988) Science 241:812-816. The fusion partner can be constructed such that it can be cleaved off such that a protein of substantially natural length is generated.

Such polypeptides may also have amino acid residues which have been chemically modified by phosphorylation, sulfonation, biotinylation, or the addition or removal of other moieties, particularly those which have molecular shapes similar to phosphate groups. In some embodiments, the modifications will be useful labeling reagents, or serve as purification targets, e.g., affinity proteins.

This invention also contemplates the use of derivatives of SOCS or WDS protein other than variations in amino acid sequence or glycosylation. Such derivatives may involve covalent or aggregative association with chemical moieties. These derivatives generally fall into the three classes: (1) salts, (2) side chain and terminal residue covalent modifications, and (3) adsorption complexes, for example with cell membranes. Such covalent or aggregative derivatives are useful as immunogens, as reagents in immunoassays, or in purification methods such as for affinity purification of proteinss or other binding proteins. For example, a SOCS or WDS protein antigen can be immobilized by covalent bonding to a solid support such as cyanogen bromide-activated SEPHAROSE, by methods which are well known in the art, or adsorbed onto polyolefin surfaces, with or without glutaraldehyde cross-linking, for use in the

assay or purification of anti-SOCS or anti-WDS protein antibodies or its respective binding partner. The SOCS or WDS protein can also be labeled with a detectable group, e.g., radioiodinated by the chloramine T procedure, covalently bound to rare earth chelates, or conjugated to another fluorescent moiety for use in diagnostic assays. Purification of SOCS or WDS proteins may be effected by immobilized antibodies or binding partner.

Isolated SOCS or WDS protein genes will allow transformation of cells lacking expression of corresponding SOCS or WDS protein, e.g., either species types or cells which lack corresponding proteins and exhibit negative background activity. Expression of transformed genes will allow isolation of antigenically pure cell lines, with defined or single specie variants. This approach will allow for more sensitive detection and discrimination of the physiological effects of SOCS or WDS binding proteins. Subcellular fragments, e.g., cytoplasts or membrane fragments, can be isolated and used.

VIII. Binding Agent:SOCS or :WDS Protein Complexes

A SOCS or WDS protein that specifically binds to or that is specifically immunoreactive with an antibody generated against a defined immunogen, such as an immunogen consisting of the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, or 16 is typically determined in an immunoassay. The immunoassay uses a polyclonal antiserum which was raised to a protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, or 16. This antiserum is selected to have low crossreactivity against other intracellular regulatory proteins and any such crossreactivity is removed by immunoabsorbtion prior to use in the immunoassay.

In order to produce antisera for use in an immunoassay, the protein of desired sequence, e.g., SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, and/or 16, is isolated as described herein. For example, recombinant protein may be produced in a mammalian cell line. An inbred strain

of mice such as Balb/c is immunized with the protein of appropriate sequence using a standard adjuvant, such as Freund's adjuvant, and a standard mouse immunization protocol (see Harlow and Lane, supra). Alternatively, a synthetic peptide, preferably near full length, derived from the sequences disclosed herein and conjugated to a carrier protein can be used an immunogen. Polyclonal sera are collected and titered against the immunogen protein in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Polyclonal antisera with a titer of 10^4 or greater are selected and tested for their cross reactivity against other intracellular proteins, using a competitive binding immunoassay such as the one described in Harlow and Lane, supra, at pages 570-573. Preferably two intracellular proteins are used in this determination in conjunction with the desired SOCS or WDS protein.

Immunoassays in the competitive binding format can be used for the crossreactivity determinations. For example, a protein of SEQ ID NO: 2 or 4 can be immobilized to a solid support. Proteins added to the assay compete with the binding of the antisera to the immobilized antigen. The ability of the above proteins to compete with the binding of the antisera to the immobilized protein is compared to the protein of SEQ ID NO: 2 or 4. The percent crossreactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% crossreactivity with each of the proteins listed above are selected and pooled. The cross-reacting antibodies are then removed from the pooled antisera by immunoabsorbtion with the above-listed proteins.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein to the immunogen protein (e.g., the SOCS14 or SOCS15 protein of SEQ ID NO: 2 and 6, or 4). In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the

immobilized protein is determined. If the amount of the second protein required is less than twice the amount of the protein, e.g., of SEQ ID NO: 2 that is required, then the second protein is said to specifically bind to an antibody generated to the immunogen.

It is understood that each of SOCS or WDS proteins are members of respective families of homologous proteins that comprise two or more genes. For a particular gene product, such as the human SOCS14 or SOCS15 protein, the term refers not only to the amino acid sequences disclosed herein, but also to other proteins that are polymorphic, allelic, non-allelic, or species variants. It is also understood that the term "SOCS14 or SOCS15 protein" includes nonnatural mutations introduced by deliberate mutation using conventional recombinant technology such as single site mutation, or by excising short sections of DNA encoding SOCS14 or SOCS15 proteins, or by substituting new amino acids, or adding new amino acids. Such minor alterations should substantially maintain the immunoidentity of the original molecule and/or its biological activity. Thus, these alterations include proteins that are specifically immunoreactive with a designated naturally occurring SOCS or WDS protein, for example, the human SOCS14 or SOCS15 protein shown in SEQ ID NO: 2 and 6, or 4 and 8. The biological properties of the altered proteins can be determined by expressing the protein in an appropriate cell line and measuring, e.g., a proliferative effect. Particular protein modifications considered minor would include conservative substitution of amino acids with similar chemical properties, as described above for the SOCS14 or SOCS15 protein as a whole. By aligning a protein optimally with the protein of SEQ ID NO: 2, 4, 6, or 8, and by using the conventional immunoassays described herein to determine immunoidentity, or by using proliferative assays, one can determine the protein compositions of the invention.

IX. Uses

The present invention provides reagents which will find use in diagnostic applications as described elsewhere herein, e.g., in the general description for developmental abnormalities, or below in the description of kits for diagnosis. Moreover, the SOCS proteins can block signaling via cytokine receptors.

SOCS or WDS nucleotides, e.g., human SOCS14 or SOCS15 DNA or RNA, may be used as a component in a forensic assay. For instance, the nucleotide sequences provided may be labeled using, e.g., ^{32}P or biotin and used to probe standard restriction fragment polymorphism blots, providing a measurable character to aid in distinguishing between individuals. Such probes may be used in well-known forensic techniques such as genetic fingerprinting. In addition, nucleotide probes made from SOCS or WDS sequences may be used in in situ assays to detect chromosomal abnormalities. For instance, rearrangements in the human chromosome encoding a SOCS14 or SOCS15 gene may be detected via well-known in situ techniques, using SOCS14 or SOCS15 probes in conjunction with other known chromosome markers.

Antibodies and other binding agents directed towards SOCS or WDS proteins or nucleic acids may be used to purify the corresponding SOCS or WDS molecule. As described in the Examples below, antibody purification of SOCS or WDS protein components is both possible and practicable. Antibodies and other binding agents may also be used in a diagnostic fashion to determine whether SOCS or WDS protein components are present in a tissue sample or cell population using well-known techniques described herein. The ability to attach a binding agent to a SOCS or WDS protein provides a means to diagnose disorders associated with SOCS or WDS protein misregulation. Antibodies and other SOCS or WDS protein binding agents may also be useful as histological markers. It is likely that specific SOCS or WDS protein expression is limited to specific tissue types. By directing a probe, such as an antibody or nucleic acid to a SOCS14 or SOCS15 protein it is possible to use the

probe to distinguish tissue and cell types in situ or in vitro.

This invention also provides reagents with significant therapeutic value. The SOCS or WDS protein (naturally occurring or recombinant), fragments thereof, and antibodies thereto, along with compounds identified as having binding affinity to a SOCS or WDS protein, are useful in the treatment of conditions associated with abnormal physiology or development, including abnormal proliferation, e.g., cancerous conditions, or degenerative conditions. Abnormal proliferation, regeneration, degeneration, and atrophy may be modulated by appropriate therapeutic treatment using the compositions provided herein. For example, a disease or disorder associated with abnormal expression or abnormal signaling by a SOCS or WDS protein is a target for an agonist or antagonist of the protein. The proteins likely play a role in regulation or development of neuronal or hematopoietic cells, e.g., lymphoid cells, which affect immunological responses.

For example, SOCS or WDS proteins likely play a role in T cell activation deficiencies in which patients develop clinical manifestations of T cell immunodeficiency such as opportunistic infections, recurrent viral or bacterial infections, diarrhea, autoimmune hemolytic anemia, lymphoid hepatitis and dermatitis, and Hodgkin lymphoma, at various stages of childhood. An excess of SOCS proteins might lead to SCID-like (severe combined immunodeficiencies) syndromes while a deficit of SOCS or WDS proteins may lead to malignant growth, for example, adult T cell leukemia/lymphoma is a disease associated with uncontrolled T-cell proliferation and is correlated at the molecular level with the presence of the IL-2 receptor (Schechter, G.P.; "Chronic Lymphocytic Leukemia" in <u>Clinical Immunology: Principles and Practice</u>, Rich (ed.) Mosby, St. Louis (Curr. ed.)). A model for adult T cell leukemia suggests that the disease may result from constitutive activation of the IL-2 receptor and its subsequent constitutive signaling cascade.

Administration of exogenous SOCS to effected T cells may modulate this disease.

Other abnormal developmental conditions are known in cell types shown to possess SOCS or WDS protein mRNA by northern blot analysis. See Berkow (ed.) The Merck Manual of Diagnosis and Therapy, Merck & Co., Rahway, N.J.; Thorn et al. Harrison's Principles of Internal Medicine, McGraw-Hill, N.Y.; and Rich (ed.) Clinical Immunology; Principles and Practice, Mosby, St. Louis (Curr. ed.). Developmental or functional abnormalities, e.g., of the neuronal or immune system, cause significant medical abnormalities and conditions which may be susceptible to prevention or treatment using compositions provided herein.

Recombinant SOCS or WDS protein or SOCS or WDS antibodies can be purified and then administered to a patient. These reagents can be combined for therapeutic use with additional active or inert ingredients, e.g., in conventional pharmaceutically acceptable carriers or diluents, e.g., immunogenic adjuvants, along with physiologically innocuous stabilizers and excipients. These combinations can be sterile filtered and placed into dosage forms as by lyophilization in dosage vials or storage in stabilized aqueous preparations. This invention also contemplates use of antibodies or binding fragments thereof, including forms which are not complement binding.

Drug screening using antibodies or fragments thereof can identify compounds having binding affinity to SOCS or WDS protein, including isolation of associated components. Subsequent biological assays can then be utilized to determine if the compound has intrinsic stimulating activity and is therefore a blocker or antagonist in that it blocks the activity of the protein. Likewise, a compound having intrinsic stimulating activity can activate the binding partner and is thus an agonist in that it simulates the activity of a SOCS or WDS protein. This invention further contemplates the therapeutic use of antibodies to SOCS or WDS protein as

antagonists. This approach should be particularly useful with other SOCS or WDS protein species variants.

Another therapeutic approach included within the invention involves direct administration of reagents or compositions by any conventional administration techniques (for example but not restricted to local injection, inhalation, or administered systemically), to the subject with an immune, allergic or trauma disorder. The reagents, formulations or compositions included within the bounds and metes of the invention may also be targeted to specific cells by any of the methods described herein. The actual dosage of reagent, formulation or composition that modulates an immune, disorder depends on many factors, including the size and health of an organism, however one of one of ordinary skill in the art can use the following teachings describing the methods and techniques for determining clinical dosages. Spilker (1984) Guide to Clinical Studies and Developing Protocols, Raven Press Books, Ltd., New York, pp. 7-13, 54-60; Spilker (1991) Guide to Clinical Trials, Raven Press, Ltd., New York, pp. 93-101; Craig and Stitzel (eds. 1986) Modern Pharmacology, 2d ed., Little, Brown and Co., Boston, pp. 127-33; Speight (ed. 1987) Avery's Drug Treatment: Principles and Practice of Clinical Pharmacology and Therapeutics, 3d ed., Williams and Wilkins, Baltimore, pp. 50-56; Tallarida, et al. (1988) Principles in General Pharmacology, Springer-Verlag, New York, pp. 18-20) to determine the appropriate dosage to use; but, generally, in the range of about between 0.5 fg/ml and 500 μg/ml inclusive final concentration are administered per day to an adult in any pharmaceutically-acceptable carrier.

The quantities of reagents necessary for effective therapy will depend upon many different factors, including means of administration, target site, physiological state of the patient, and other medicants administered. Thus, treatment dosages should be titrated to optimize safety and efficacy. Typically, dosages used in vitro may provide useful guidance in the amounts useful for in situ administration of these reagents.

Animal testing of effective doses for treatment of particular disorders will provide further predictive indication of human dosage. Various considerations are described, e.g., in Gilman, et al. (eds.) (1990) Goodman and Gilman's: The Pharmacological Bases of Therapeutics (8th ed.) Pergamon Press; and (1990) Remington's Pharmaceutical Sciences (17th ed.) Mack Publishing Co., Easton, PA. Methods for administration are discussed therein and below, e.g., for oral, intravenous, intraperitoneal, or intramuscular administration, transdermal diffusion, and others. Pharmaceutically acceptable carriers will include water, saline, buffers, and other compounds described, e.g., in the Merck Index, Merck & Co., Rahway, NJ. Dosage ranges would ordinarily be expected to be in amounts lower than 1 mM concentrations, typically less than about 10 μM concentrations, usually less than about 100 nM, preferably less than about 10 pM (picomolar), and most preferably less than about 1 fM (femtomolar), with an appropriate carrier. Slow release formulations, or a slow release apparatus will often be utilized for continuous administration.

SOCS or WDS protein, fragments thereof, and antibodies to it or its fragments, antagonists, and agonists, may be administered directly to the host to be treated or, depending on the size of the compounds, it may be desirable to conjugate them to carrier proteins such as ovalbumin or serum albumin prior to their administration. Therapeutic formulations may be administered in any conventional dosage formulation. While it is possible for the active ingredient to be administered alone, it is preferable to present it as a pharmaceutical formulation. Formulations typically comprise at least one active ingredient, as defined above, together with one or more acceptable carriers thereof. Each carrier should be both pharmaceutically and physiologically acceptable in the sense of being compatible with the other ingredients and not injurious to the patient. Formulations include those suitable for oral, rectal, nasal, or parenteral (including

subcutaneous, intramuscular, intravenous and intradermal) administration. The formulations may conveniently be presented in unit dosage form and may be prepared by any methods well known in the art of pharmacy. See, e.g., Gilman, et al. (eds.) (1990) Goodman and Gilman's: The Pharmacological Bases of Therapeutics (8th ed.) Pergamon Press; and (1990) Remington's Pharmaceutical Sciences (17th ed.) Mack Publishing Co., Easton, PA; Avis, et al. (eds.) (1993) Pharmaceutical Dosage Forms: Parenteral Medications Dekker, NY; Lieberman, et al. (eds.) (1990) Pharmaceutical Dosgae Forms: Tablets Dekker, NY; and Lieberman, et al. (eds.) (1990) Pharmaceutical Dosage Forms: Disperse Systems Dekker, NY. The therapy of this invention may be combined with or used in association with other therapeutic agents.

Both the naturally occurring and the recombinant forms of the SOCS or WDS proteins of this invention are particularly useful in kits and assay methods which are capable of screening compounds for binding activity to the proteins. Several methods of automating assays have been developed in recent years so as to permit screening of tens of thousands of compounds in a short period. See, e.g., Fodor, et al. (1991) Science 251:767-773, and other descriptions of chemical diversity libraries, which describe means for testing of binding affinity by a plurality of compounds. The development of suitable assays can be greatly facilitated by the availability of large amounts of purified, soluble SOCS or WDS protein as provided by this invention.

For example, antagonists can normally be found once the protein has been structurally defined. Testing of potential protein analogs is now possible upon the development of highly automated assay methods using a purified binding partner. In particular, new agonists and antagonists will be discovered by using screening techniques described herein. Of particular importance are compounds found to have a combined binding affinity for multiple SOCS or WDS protein binding components, e.g., compounds which can serve as antagonists for species variants of a SOCS or WDS protein.

This invention is particularly useful for screening compounds by using recombinant protein in a variety of drug screening techniques. The advantages of using a recombinant protein in screening for specific binding partners include: (a) improved renewable source of the SOCS or WDS protein from a specific source; (b) potentially greater number of binding partners per cell giving better signal to noise ratio in assays; and (c) species variant specificity (theoretically giving greater biological and disease specificity).

One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant DNA molecules expressing a SOCS or WDS protein binding counterpart. Cells may be isolated which express a binding counterpart in isolation from any others. Such cells, either in viable or fixed form, can be used for standard protein binding assays. See also, Parce, et al. (1989) Science 246:243-247; and Owicki, et al. (1990) Proc. Nat'l Acad. Sci. USA 87:4007-4011, which describe sensitive methods to detect cellular responses. Competitive assays are particularly useful, where the cells (source of SOCS14 or SOCS15 protein) are contacted and incubated with a labeled binding partner or antibody having known binding affinity to the protein, such as ^{125}I-antibody, and a test sample whose binding affinity to the binding composition is being measured. The bound and free labeled binding compositions are then separated to assess the degree of protein binding. The amount of test compound bound is inversely proportional to the amount of labeled binding partner binding to the known source. Any one of numerous techniques can be used to separate bound from free protein to assess the degree of protein binding. This separation step could typically involve a procedure such as adhesion to filters followed by washing, adhesion to plastic followed by washing, or centrifugation of the cell membranes. Viable cells could also be used to screen for the effects of drugs on SOCS or WDS protein mediated functions, e.g., second messenger levels, i.e., cell proliferation; inositol phosphate pool changes, transcription using a luciferase-type assay; and

others. Some detection methods allow for elimination of a separation step, e.g., a proximity sensitive detection system.

Another method utilizes membranes from transformed eukaryotic or prokaryotic host cells as the source of a SOCS or WDS protein. These cells are stably transformed with DNA vectors directing the expression of a SOCS or WDS protein, e.g., an engineered membrane bound form. Essentially, the membranes would be prepared from the cells and used in a protein binding assay such as the competitive assay set forth above.

Still another approach is to use solubilized, unpurified or solubilized, purified SOCS or WDS protein from transformed eukaryotic or prokaryotic host cells. This allows for a "molecular" binding assay with the advantages of increased specificity, the ability to automate, and high drug test throughput.

Another technique for drug screening involves an approach which provides high throughput screening for compounds having suitable binding affinity to a SOCS or WDS protein antibody and is described in detail in Geysen, European Patent Application 84/03564, published on September 13, 1984. First, large numbers of different small peptide test compounds are synthesized on a solid substrate, e.g., plastic pins or some other appropriate surface, see Fodor, et al., supra. Then all the pins are reacted with solubilized, unpurified or solubilized, purified SOCS or WDS protein antibody, and washed. The next step involves detecting bound SOCS or WDS protein antibody.

Rational drug design may also be based upon structural studies of the molecular shapes of the SOCS or WDS protein and other effectors or analogs. See, e.g., Methods in Enzymology vols 202 and 203. Effectors may be other proteins which mediate other functions in response to protein binding, or other proteins which normally interact with the binding partner. One means for determining which sites interact with specific other proteins is a physical structure determination, e.g., x-ray crystallography or 2 dimensional NMR techniques.

These will provide guidance as to which amino acid residues form molecular contact regions. For a detailed description of protein structural determination, see, e.g., Blundell and Johnson (1976) Protein Crystallography Academic Press, NY.

A purified SOCS or WDS protein can be coated directly onto plates for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to these proteins can be used as capture antibodies to immobilize the respective protein on the solid phase.

X. Kits

This invention also contemplates use of SOCS or WDS proteins, fragments thereof, peptides, and their fusion products in a variety of diagnostic kits and methods for detecting the presence of SOCS or WDS protein or a binding partner. Typically the kit will have a compartment containing either a defined SOCS or WDS protein peptide or gene segment or a reagent which recognizes one or the other, e.g., binding partner fragments or antibodies.

A kit for determining the binding affinity of a test compound to a SOCS or WDS protein would typically comprise a test compound; a labeled compound, e.g., a binding agent or antibody having known binding affinity for the SOCS or WDS protein; a source of SOCS or WDS protein (naturally occurring or recombinant); and a means for separating bound from free labeled compound, such as a solid phase for immobilizing the SOCS or WDS protein. Once compounds are screened, those having suitable binding affinity to the SOCS or WDS protein can be evaluated in suitable biological assays, as are well known in the art, to determine whether they act as agonists or antagonists to the binding partner. The availability of recombinant SOCS or WDS protein polypeptides also provide well defined standards for calibrating such assays.

A preferred kit for determining the concentration of, for example, a SOCS or WDS protein in a sample would

typically comprise a labeled compound, e.g., binding partner or antibody, having known binding affinity for the SOCS or WDS protein, a source of SOCS or WDS protein (naturally occurring or recombinant), and a means for separating the bound from free labeled compound, for example, a solid phase for immobilizing the SOCS or WDS protein. Compartments containing reagents, and instructions, will normally be provided.

Antibodies, including antigen binding fragments, specific for the SOCS or WDS protein or fragments thereof are useful in diagnostic applications to detect the presence of elevated levels of SOCS or WDS protein and/or its fragments. Such diagnostic assays can employ lysates, live cells, fixed cells, immunofluorescence, cell cultures, body fluids, and further can involve the detection of antigens related to the protein in serum, or the like. Diagnostic assays may be homogeneous (without a separation step between free reagent and antigen-SOCS or -WDS protein complex) or heterogeneous (with a separation step). Various commercial assays exist, such as radioimmunoassay (RIA), enzyme-linked immunosorbentassay (ELISA), enzyme immunoassay (EIA), enzyme-multiplied immunoassay technique (EMIT), substrate-labeled fluorescent immunoassay (SLFIA), and the like. For example, unlabeled antibodies can be employed by using a second antibody which is labeled and which recognizes the antibody to a SOCS or WDS protein or to a particular fragment thereof. Similar assays have also been extensively discussed in the literature. See, e.g., Harlow and Lane (1988) Antibodies: A Laboratory Manual, CSH Press, NY; Chan (ed.) (1987) Immunoassay: A Practical Guide Academic Press, Orlando, FL; Price and Newman (eds.) (1991) Principles and Practice of Immunoassay Stockton Press, NY; and Ngo (ed.) (1988) Nonisotopic Immunoassay Plenum Press, NY.

Anti-idiotypic antibodies may have similar use to diagnose presence of antibodies against a SOCS or WDS protein, as such may be diagnostic of various abnormal states. For example, overproduction of SOCS or WDS protein may result in production of various immunological

or other medical reactions which may be diagnostic of abnormal physiological states, e.g., in cell growth, activation, or differentiation.

Frequently, the reagents for diagnostic assays are supplied in kits, so as to optimize the sensitivity of the assay. For the subject invention, depending upon the nature of the assay, the protocol, and the label, either labeled or unlabeled antibody or binding partner, or labeled SOCS or WDS protein is provided. This is usually in conjunction with other additives, such as buffers, stabilizers, materials necessary for signal production such as substrates for enzymes, and the like. Preferably, the kit will also contain instructions for proper use and disposal of the contents after use. Typically the kit has compartments for each useful reagent. Desirably, the reagents are provided as a dry lyophilized powder, where the reagents may be reconstituted in an aqueous medium providing appropriate concentrations of reagents for performing the assay.

Many of the aforementioned constituents of the drug screening and the diagnostic assays may be used without modification, or may be modified in a variety of ways. For example, labeling may be achieved by covalently or non-covalently joining a moiety which directly or indirectly provides a detectable signal. In any of these assays, the protein, test compound, SOCS or WDS protein, or antibodies thereto can be labeled either directly or indirectly. Possibilities for direct labeling include label groups: radiolabels such as ^{125}I, enzymes (U.S. Pat. No. 3,645,090) such as peroxidase and alkaline phosphatase, and fluorescent labels (U.S. Pat. No. 3,940,475) capable of monitoring the change in fluorescence intensity, wavelength shift, or fluorescence polarization. Possibilities for indirect labeling include biotinylation of one constituent followed by binding to avidin coupled to one of the above label groups.

There are also numerous methods of separating the bound from the free protein, or alternatively the bound from the free test compound. The SOCS or WDS protein can

be immobilized on various matrices followed by washing. Suitable matrices include plastic such as an ELISA plate, filters, and beads. Methods of immobilizing the SOCS or WDS protein to a matrix include, without limitation, direct adhesion to plastic, use of a capture antibody, chemical coupling, and biotin-avidin. The last step in this approach involves the precipitation of protein/binding partner or antigen/antibody complex by any of several methods including those utilizing, e.g., an organic solvent such as polyethylene glycol or a salt such as ammonium sulfate. Other suitable separation techniques include, without limitation, the fluorescein antibody magnetizable particle method described in Rattle, et al. (1984) Clin. Chem. 30:1457-1461, and the double antibody magnetic particle separation as described in U.S. Pat. No. 4,659,678.

Methods for linking proteins or their fragments to the various labels have been extensively reported in the literature and do not require detailed discussion here. Many of the techniques involve the use of activated carboxyl groups either through the use of carbodiimide or active esters to form peptide bonds, the formation of thioethers by reaction of a mercapto group with an activated halogen such as chloroacetyl, or an activated olefin such as maleimide, for linkage, or the like. Fusion proteins will also find use in these applications.

Another diagnostic aspect of this invention involves use of oligonucleotide or polynucleotide sequences taken from the sequence of a SOCS or WDS protein. These sequences can be used as probes for detecting levels of the SOCS or WDS protein message in samples from natural sources, or patients suspected of having an abnormal condition, e.g., cancer or developmental problem. The preparation of both RNA and DNA nucleotide sequences, the labeling of the sequences, and the preferred size of the sequences has received ample description and discussion in the literature. Normally an oligonucleotide probe should have at least about 14 nucleotides, usually at least about 18 nucleotides, and the polynucleotide probes may be up to several kilobases. Various labels may be

employed, most commonly radionuclides, particularly ^{32}P. However, other techniques may also be employed, such as using biotin modified nucleotides for introduction into a polynucleotide. The biotin then serves as the site for binding to avidin or antibodies, which may be labeled with a wide variety of labels, such as radionuclides, fluorophores, enzymes, or the like. Alternatively, antibodies may be employed which can recognize specific duplexes, including DNA duplexes, RNA duplexes, DNA-RNA hybrid duplexes, or DNA-protein duplexes. The antibodies in turn may be labeled and the assay carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected. The use of probes to the novel anti-sense RNA may be carried out using many conventional techniques such as nucleic acid hybridization, plus and minus screening, recombinational probing, hybrid released translation (HRT), and hybrid arrested translation (HART). This also includes amplification techniques such as polymerase chain reaction (PCR).

Diagnostic kits which also test for the qualitative or quantitative presence of other markers are also contemplated. Diagnosis or prognosis may depend on the combination of multiple indications used as markers. Thus, kits may test for combinations of markers. See, e.g., Viallet, et al. (1989) Progress in Growth Factor Res. 1:89-97.

The broad scope of this invention is best understood with reference to the following examples, which are not intended to limit the invention to specific embodiments.

EXAMPLES

I. General Methods

Many of the standard methods below are described or referenced, e.g., in Maniatis, et al. (Cur. ed..) Molecular Cloning, A Laboratory Manual Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY; Sambrook, et al. (1989) Molecular Cloning: A Laboratory Manual (2d ed.) Vols. 1-3, CSH Press, NY; Ausubel, et al., Biology Greene Publishing Associates, Brooklyn, NY; or Ausubel, et al. (1987 and Supplements) Current Protocols in Molecular Biology Wiley/Greene, NY; Innis, et al. (eds.) (1990) PCR Protocols: A Guide to Methods and Applications Academic Press, NY. Methods for protein purification include such methods as ammonium sulfate precipitation, column chromatography, electrophoresis, centrifugation, crystallization, and others. See, e.g., Ausubel, et al. (1987 and periodic supplements); Deutscher (1990) "Guide to Protein Purification," Methods in Enzymology vol. 182, and other volumes in this series; Coligan, et al. (1995 and supplements) Current Protocols in Protein Science John Wiley and Sons, New York, NY; P. Matsudaira (ed.) (1993) A Practical Guide to Protein and Peptide Purification for Microsequencing, Academic Press, San Diego, CA; and manufacturer's literature on use of protein purification products, e.g., Pharmacia, Piscataway, NJ, or Bio-Rad, Richmond, CA. Combination with recombinant techniques allow fusion to appropriate segments (epitope tags), e.g., to a FLAG sequence or an equivalent which can be fused, e.g., via a protease-removable sequence. See, e.g., Hochuli (1989) Chemische Industrie 12:69-70; Hochuli (1990) "Purification of Recombinant Proteins with Metal Chelate Absorbent" in Setlow (ed.) Genetic Engineering, Principle and Methods 12:87-98, Plenum Press, NY; and Crowe, et al. (1992) QIAexpress: The High Level Expression & Protein Purification System QUIAGEN, Inc., Chatsworth, CA.

Standard immunological techniques are described, e.g., in Hertzenberg, et al. (eds. 1996) Weir's Hanbook

of Experimental Immunology vols 1-4, Blackwell Science; Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; and Methods in Enzymology volumes. 70, 73, 74, 84, 92, 93, 108, 116, 121, 132, 150, 162, and 163. Assays for neural cell biological activities are described, e.g., in Wouterlood (ed. 1995) Neuroscience Protocols modules 10, Elsevier; Methods in Neurosciences Academic Press; and Neuromethods Humana Press, Totowa, NJ. Methodology of developmental systems is described, e.g., in Meisami (ed.) Handbook of Human Growth and Developmental Biology CRC Press; and Chrispeels (ed.) Molecular Techniques and Approaches in Developmental Biology Interscience.

FACS analyses are described in Melamed, et al. (1990) Flow Cytometry and Sorting Wiley-Liss, Inc., New York, NY; Shapiro (1988) Practical Flow Cytometry Liss, New York, NY; and Robinson, et al. (1993) Handbook of Flow Cytometry Methods Wiley-Liss, New York, NY.

II. Isolation of full length SOCS or WDS clones

Standard methods are used to isolate full length genes. A cDNA library from an appropriate, e.g., human, cell, preferably a STAT containing cell type. The appropriate sequence is selected, and hybridization at high stringency conditions is performed to find a full length corresponding gene. It is noted that the mouse and human protein sequences are virtually identical.

III. Isolation of primate SOCS14 or SOCS15 clones

The full length, or appropriate fragments, of human genes are used to isolate a corresponding monkey or other primate gene. Preferably a full length coding sequence is used for hybridization. Similar source materials as indicated above are used to isolate natural genes, including genetic, polymorphic, allelic, or strain variants. Other species variants are also isolated using similar methods.

IV. Isolation of an avian SOCS14 or SOCS15 clone

An appropriate avian source is selected as above. Similar methods are utilized to isolate a species variant, though the level of similarity will typically be lower for avian protein as compared to a human to mouse sequence.

V. Expression; purification; characterization

Proteins of interest are immunoprecipitated and affinity purified as described above, e.g., from a natural or recombinant source.

Alternatively, with an appropriate clone from above, the coding sequence is inserted into an appropriate expression vector. This may be in a vector specifically selected for a prokaryote, yeast, insect, or higher vertebrate, e.g., mammalian expression system. Standard methods are applied to produce the gene product, preferably as a soluble secreted molecule, but will, in certain instances, also be made as an intracellular protein. Intracellular proteins typically require cell lysis to recover the protein, and insoluble inclusion bodies are a common starting material for further purificiation.

With a clone encoding a vertebrate SOCS14 or SOCS15 protein, recombinant production means are used, although natural forms may be purified from appropriate sources. The protein product is purified by standard methods of protein purification, in certain cases, e.g., coupled with immunoaffinity methods. Immunoaffinity methods are used either as a purification step, as described above, or as a detection assay to determine the separation properties of the protein.

Preferably, the protein is secreted into the medium, and the soluble product is purified from the medium in a soluble form. Alternatively, as described above, inclusion bodies from prokaryotic expression systems are a useful source of material. Typically, the insoluble protein is solubilized from the inclusion bodies and refolded using standard methods. Purification methods are developed as described above.

The product of the purification method described above is characterized to determine many structural features. Standard physical methods are applied, e.g., amino acid analysis and protein sequencing. The resulting protein is subjected to CD spectroscopy and other spectroscopic methods, e.g., NMR, ESR, mass spectroscopy, etc. The product is characterized to determine its molecular form and size, e.g., using gel chromatography and similar techniques. Understanding of the chromatographic properties will lead to more gentle or efficient purification methods.

Prediction of glycosylation sites may be made, e.g., as reported in Hansen, et al. (1995) Biochem. J. 308:801-813. However, as intracellular proteins, they are unlikely to be normally glycosylated.

The purified protein is also be used to identify other binding partners of SOCS or WDS as described, e.g., in Fields and Song (1989) Nature 340:245-246.

VI. Preparation of antibodies against vertebrate SOCS or WDS

With protein produced, as above, animals are immunized to produce antibodies. Polyclonal antiserum is raised using non-purified antigen, though the resulting serum will exhibit higher background levels. Preferably, the antigen is purified using standard protein purification techniques, including, e.g., affinity chromatography using polyclonal serum indicated above. Presence of specific antibodies is detected using defined synthetic peptide fragments.

Polyclonal serum is raised against a purified antigen, purified as indicated above, or using, e.g., a plurality of, synthetic peptides. A series of overlapping synthetic peptides which encompass all of the full length sequence, if presented to an animal, will produce serum recognizing most linear epitopes on the protein. Such an antiserum is used to affinity purify protein, which is, in turn, used to introduce intact full length protein into another animal to produce another antiserum preparation.

Similar techniques are used to generate induce monoclonal antibodies to either unpurified antigen, or, preferably, purified antigen.

VII. Cellular and tissue distribution

Distribution of the protein or gene products are determined, e.g., using immunohistochemistry with an antibody reagent, as produced above, by Western blotting of cell lysates, or by screening for nucleic acids encoding the respective protein. Either hybridization or PCR methods are used to detect DNA, cDNA, or message content. Histochemistry allows determination of the specific cell types within a tissue which express higher or lower levels of message or DNA. Antibody techniques are useful to quantitate protein in a biological sample, including a liquid or tissue sample. Immunoassays are developed to quantitate protein. Also FACS analysis may be used to evaluate expression in a cell population. Appropriate tissue samples or cell types are isolated and prepared for such detection. Commercial tissue blots are available, e.g., from Clontech (Mountain View, CA). Alternatively, cDNA library Southern blots can be analyzed.

VIII. STAT interference by SOCS or WDS proteins

Standard methods for testing the biological activity of the SOCS gene products in STAT signaling are described, e.g., in Starr, et al. (1997) Nature 387:917-921; Endo, et al. (1997) Nature 387:921-924; and Naka, et al. Nature 387:924-929. Alternatively, JAK/STATs are necessary for signal transduction. This assay is performed as described, e.g., in Ho, et al. (1995) Mol. Cell. Biol. 15:5043-5-53, and blockage with these gene products may be tested.

In particular, the STAT5 dependent signaling in response to IL-2 is inhibited by the SOCS family member SOCS3.

IX. Antagonists of SOCS function

The inhibition of SOCS function may be effected by inhibitors of the specific interaction of these gene products and their respective STAT molecules. With the information on the specificity of pairings between these SOCS and respective STAT family members, compound libraries may be screened for blockage of such interactions. Thus, inhibitory action of the SOCS may be blocked with small molecule drug candidates.

Methods of using gene therapy are described, e.g., in Goodnow (1992) "Transgenic Animals" in Roitt (ed.) Encyclopedia of Immunology, Academic Press, San Diego, pp. 1502-1504; Travis (1992) Science 256:1392-1394; Kuhn, et al. (1991) Science 254:707-710; Capecchi (1989) Science 244:1288; Robertson (1987)(ed.) Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, IRL Press, Oxford; and Rosenberg (1992) J. Clinical Oncology 10:180-199. Also included is the use of antisense RNA in gene therapy to block expression of the target gene, or proper splicing of gene transcripts.

X. Comparison of various SOCS embodiments

Tables 1 and 2 show comparison of various SOCS or WDS embodiments. Table 1 shows comparisons of the relevant portions of the gene products, particularly in the region of SOCS14 from Met168 to Leu293.

Table 2 shows alignment of the WDS "SOCSBOX protein" with a consensus of the mouse and human SOCS15 (WDS11) protein sequences, which are identical. See GenBank Accession numbers U88325; U88326; U88327; U88328; AB000676; AB000677; AB000710. This is aligned with the new WDS12, SEQ ID NO: 16.

Table 1: Comparison of various SOCS family members. mCIS is SEQ ID NO: 15; hSOCS1 IS SEQ ID NO: 16; mSOCS1 is SEQ ID NO: 17; hSOCS2 is SEQ ID NO: 18; hSOCS3 is SEQ ID NO: 19; mSOCS3 is SEQ ID NO: 20; and human SOCS16 is SEQ ID NO: 21.

```
mCIS
hSOCS1
mSOCS1
hSOCS2
hSOCS3
mSOCS3
hSOCS14                                      MEVRVKALVHSSS
mSOCS17                       AELGEIR-----------PESAQKKLPLRKA
hSOCS18   MDKVGKMWNNLKYRCQNLFSHEGGSRNENVEMNPNRCPSVKEKSISLGEA
hSOCS19                       ERGLETNSCSEEELSSPGRGGGGGGRLLLQ

mCIS
hSOCS1
mSOCS1
hSOCS2                     ALSPAATLTAWPADSARRGP-------------
hSOCS3
mSOCS3
hSOCS14   PSPALNGVRKDFHDLQSETTCQEQANSLKSSASHNGDLHLHLDEHVPVVI
mSOCS17   EN-----------TIFITLEIVKNLFKMAENNSKNVDVRPKTSRSRSAD-
hSOCS18   APQQESSPLRENVALQLGLSPSKTFSRRNQNCAAEIPQVVEISIEKDSDS
hSOCS19   PPGPELPPVPFPLQDLVPLGRLSRGEQQQQQQQPPPPPPPPGPLRPLAG

mCIS
hSOCS1
mSOCS1
hSOCS2    --------------------------------------------------
hSOCS3
mSOCS3
hSOCS14   G-----------------LMPQDYIQYTVPLDEGMYPLEGSRS-------
mSOCS17   ---------RKD----GYVWSGKK-LSWSKKSESCSESEAKKG-------
hSOCS18   GATPGTRLARRDSYSRHAPWGGKKKHSCSTKTQSSLDTEKKFGRTRSGLQ
hSOCS19   -----------------PSRKGSFKIRLSRLFRTKSCNGGSGG-------

mCIS                                                 MVLCVQG
hSOCS1
mSOCS1
hSOCS2    -----GCTASGYPVPAARA-PAAGDQWVT--AAARDFVIR--PPGSGEKE
hSOCS3
mSOCS3
hSOCS14   -----YCLDSSSPMEVSAVPPQVGGRAFPEDESQVDQDLVVAPEIFVDQS
mSOCS17   ---QLSCSSIELDLDHSCG-HRFLGRSLK--QKLQDAVGQCFPIKNCSGR
hSOCS18   RRERRYGVSSMQDMDSVSS-RAVGSRSLR--QRLQDTVGLCFPMRTYSKQ
hSOCS19   ----GDGTGKRPSGELAAS-AASLTDMGG--SAGRELDAGRKPKLTRTQS
```

Table 1 (continued):

```
mCIS        SCPLLAVEQIGRR-PLWAQSLELPGPA------------MQPLPTGA---
hSOCS1                      MVAHNQVAADN-----------AVSTAAEPR---
mSOCS1                      MVARNQVAADN-----------AISPAAEPR---
hSOCS2      PHPFSLCHHFGHPAGLVLGFALTSRKD--------ANPSLTPARAAT---
hSOCS3                      MVTHSKFPAAG-----------MSRPLDTSL---
mSOCS3                      MVTHSKFPAAG-----------MSRPLDTSL---
hSOCS14     VNGLLIGTTGVMLQSPRAGHDDVPPLS----------PLLPPMQNNQ---
mSOCS17     HSPGLPSKRKIHISELMLDXCXFPPRSDLAFRWHFIKRHTVPMSPNS---
hSOCS18     SKPLFSNKRKIHLSELMLEKCPFPAGSDLAQKWHLIKQHTAPVSPHSTFF
hSOCS19     AFSPVSFSPLFTGETVSLVDVDISQRG-----------LTSPHPPTP---


mCIS        --------------------------------------------------
hSOCS1      -----------------RRPE----------PSSSSSSS--------PAA
mSOCS1      -----------------RRSE----------PSSSSSSSS-------PAA
hSOCS2      --------------CLCRGD-----------PS------------LMTLR
hSOCS3      -----------------R--------------------------------
mSOCS3      -----------------R--------------------------------
hSOCS14     ---------------IQRNFS--------------------------GLT
mSOCS17     ----DEWVSADLSERKLRDAQLKRRNTEDDIPCFSHTNGQPCVITANSAS
hSOCS18     DTFDPSLVSTEDEEDRLRERRRLSIEEGVDPPPNAQIHTFEATAQVNPLF
hSOCS19     -------------PPPPRRSLSLLDDISGTLPTSVLVAPMGSSLQSFPLP


mCIS        -FPEEVTEETPVQAENE----------------PKVLDP-----------
hSOCS1      PARPRPCPAVPAPAPGD--------------THFRTFRS-----------
mSOCS1      PVRPRPCPAVPAPAPGD--------------THFRTFRS-----------
hSOCS2      CLEPSGNGGEGTRSQWG--------------TAGSAEEP-----------
hSOCS3      ---------------------------------LKTFSS-----------
mSOCS3      ---------------------------------LKTFSS-----------
hSOCS14     GTEAHVAESMRCHLNFD---------PNSAPGVARVYDSVQ---------
mSOCS17     CTGGHITGSMMNLVTNN-SIEDSDMDSEDEIITLCTSSRKRNKPR--WEM
hSOCS18     KLGPKLAPGMTEISGDSSAIPQANCDSEEDTTTLCLQSR-RQKQRQISGD
hSOCS19     PPPPPHAPDAFPRIAPIR-----------AAESLHSQPP-----------


mCIS        -------------EGDLLCIAKTFSYLRES---GWYWGSITASEARQHLQ
hSOCS1      -------------HADYRRITRASALLDAC---GFYWGPLSVHGAHERLR
mSOCS1      -------------HSDYRRITRTSALLDAC---GFYWGPLSVHGAHERLR
hSOCS2      -------------SPQAARLAKALRELGQT---GWYWGSMTVNEAKEKLK
hSOCS3      -------------KSEYQLVVNAVRKLQES---GFYWSAVTGGEANLLLS
mSOCS3      -------------KSEYQLVVNAVRKLQES---GFYWSAVTGGEANLLLS
hSOCS14     -------------SSGPMVVTSLTEELKKLAKQGWYWGPITRWEAEGKLA
mSOCS17     EEEILQLEAPPKFHTQIDYVHCLVPDLLQISNNPCYWGVMDKYAAEALLE
hSOCS18     SHTHVSRQGAWKVHTQIDYIHCLVPDLLQITGNPCYWGVMDRYEAEALSE
hSOCS19     -----QHLQCPLYRPDSSSFAASLRELEKC---GWYWGPMNWEDAEMKLK
                                      *        **  .    *
```

Table 1 (continued):

```
mCIS      KMPEGTFLVRDST-HPSYLFTLSVKTTRGPTNVRIEYADSSFRLDSNCLS
hSOCS1    AEPVGTFLVRDSR-QRNCFFALSVKMASGPTSIRVHFQAGRFHLDGS-R-
mSOCS1    AEPVGTFLVRDSR-QRNCFFALSVKMASGPTSIRVHFQAGRFHLDGS-R-
hSOCS2    EAPEGTFLIRDSS-HSDYLLTISVKTSAGPTNLRIEYQDGKFRLDSIICV
hSOCS3    AEPAGTFLIRDSSDQR-HFFALSVKTQSGTKNLRIQCEGGSFSLQSDPRS
mSOCS3    AEPAGTFLIRDSSDQR-HFFTLSVKTQSGTKNLRIQCEGGSFSLQSDPRS
hSOCS14   NVPDGSFLVRDSS-DDRYLLSLSFRSHGKTLHTRIEHSNGRFSFYEQPD-
mSOCS17   GKPEGTFLLRDSA-QEDYLFSVSFRRYSRSLHARIEQWNHNFSFDAHDP-
hSOCS18   GKPEGTFLLRDSA-QEDYLFSVSSAATTGSLHARIEQWNHNFSFDAHDP-
hSOCS19   GKPDGSFLVRDSS-DPRYILSLSFRSQGITHHTRMEHYRGTFSLWCHPKF
            * *.**.***        ..*          *.      *

mCIS      RP-RILAFPDVVSLVQHYVASCAADTRSDSPDPAPTPALPMSKQDAPSDS
hSOCS1    -----ESFDCLFELLEHYVAAP-----------------------RRMLG
mSOCS1    -----ETFDCLFELLEHYVAAP-----------------------RRMLG
hSOCS2    KS-KLKQFDSVVHLIDYYVQMCKDK---------------RTGPEAPRNG
hSOCS3    TQ-PVPRFDCVLKLVYHYMPPPGAPSFP-SPPTEPSSEVPEQPSAQPLPG
mSOCS3    TQ-PVPRFDCVLKLVHHYMPPPGTPSFS-LPPTEPSSEVPEQPPAQALPG
hSOCS14   ----VERTYSIVDLIEHSIQGLENG------------AFCYSRSRLPGSA
mSOCS17   ---CVFHSPDITGLLEHYKDPSA------------------CMFFEPLLS
hSOCS18   ---CVFHSSTVTGLLEHYKDPSS------------------CMFFEPLLT
hSOCS19   EDRCQSVVEFIKRAIMHSKNGK---------------FLYFLRSRVPGLP
                        .   .

mCIS      VLPIPVATAVHLKLVQPFVRRSS----ARSLQHLCRLVINRLVA---DVD
hSOCS1    ---------------APLRQRR-----VRPLQELCRQRIVATVG-RENLA
mSOCS1    ---------------APLRQRR-----VRPLQELCRQRIVAAVG-RENLA
hSOCS2    --------------TVHLYLTKPLYTSAPSLQHLCRLTINKCTG---AIW
hSOCS3    SPPRRAYYIYSGGEKIPLVLSRPLSSNVATLQHLCRKTVNGHLDSYEKVT
mSOCS3    STPKRAYYIYSGGEKIPLVLSRPLSSNVATLQHLCRKTVNGHLDSYEKVT
hSOCS14   TYP------------VRLTNPVSRFMQVRSLQYLCRFVIRQYTR-IDLIQ
mSOCS17   ---------------TPLIRTFP-----FSLQHICRTVICNCTT-YDGID
hSOCS18   ---------------ISLNRTFP-----FSLQYICRAVICRCTT-YDGID
hSOCS19   PTP------------VQLLYPVSRFSNVKSLQHLCRFRIRQLVR-IDHIP
                                        ** .**  .         .

mCIS      CLPLPRRMADYLRQYPFQL
hSOCS1    RIPLNPVLRDYLSSFPFQI
mSOCS1    RIPLNPVLRDYLSSFPFQI
hSOCS2    GLPLPTRLKDYLEEYKFQV
hSOCS3    QLPG-P-IREFLDQYDAPL
mSOCS3    QLPG-P-IREFLDQYDAPL
hSOCS14   KLPLPNKMKDYLQEKHY
mSOCS17   ALPIPSPMKLYLKEYHYKSKVRLLRIDVPEQQ
hSOCS18   GLPLPSMLQDFLKEYHYKQKVRVRWLEREPVKAK
hSOCS19   DLPLPKPLISYIRKFYYYDPQEEVYLSLKEAQLISKQKQEVEPST
           .*    . ..
```

Table 2: Comparison of the WDS family members; WDS11 (SOCS15) and WDS12.

```
WDS12-          MLNIILIKFSSFSIRCAILSSVCLNEAITFAFLLQVFLWNMDKYTMIRKL
WDS11(socs15)                MLCSAAG------------EKSVFLWSMRSYTLIRKL
                             . *.                  ****  *  **.****

WDS12-          EGHHHDVVACDFSPDGALLATASYDTRVYIWDPHNGDILMEFGHLFPPPT
WDS11(socs15)   EGHQSSVVSCDFSPDSALLVTASYDTSVIMWDPYTGERLRSLHHTQLEPT
                ***.  **.****** *** ****** * .*** .*. *    *    **

WDS12-          PIFAGGANDRWVRSVSFSHDGLHVASLADDKMVRFWRIDEDYPVQVAPLS
WDS11(socs15)   MDDSD-VHMSSLRSVCFSPEGLYLATVADDRLLRIWALELKAPVAFAPMT
                   .   .   .***.** .** .*..***...* * ..   **  **..

WDS12-          NGLCCAFSTDGSVLAAGTHDGSVYFWATPRQVPSLQHLCRMSIRRVMPTQ
WDS11(socs15)   NGLCCTFFPHG-GIATGTRDGHVQFWTAPRVLSSLKHLCRKALRSFLTTY
                *****.*   *  .*.**.** * **..** . **.****  ..*  . *

WDS12-          EVQELPIPSKLLEFLSYRI  219
WDS11(socs15)   QVLALPIPKKMKEFLTYRTF 193
                .*  **** *. ***.**
```

All references cited herein are incorporated herein by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference in its entirety for all purposes.

Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. The specific embodiments described herein are offered by way of example only, and the invention is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

WHAT IS CLAIMED IS:

1. An isolated or recombinant polypeptide comprising:
   a) at least 17 contiguous amino acids from the coding portion of SEQ ID NO: 2 or 6;
   b) at least 17 contiguous amino acids from the coding portion of SEQ ID NO: 4 or 8;
   c) at least 17 contiguous amino acids from the coding portion of SEQ ID NO: 10;
   d) at least 17 contiguous amino acids from the coding portion of SEQ ID NO: 12;
   e) at least 17 contiguous amino acids from the coding portion of SEQ ID NO: 14; or
   f) at least 17 contiguous amino acids from the coding portion of SEQ ID NO: 16.

2. The polypeptide of claim 1, comprising the amino acid sequence of:
   a) a SOCS14 of SEQ ID NO: 2 or 6;
   b) a SOCS15 (WDS11) of SEQ ID NO: 4 or 8;
   c) a SOCS17 of SEQ ID NO: 10;
   d) a SOCS18 of SEQ ID NO: 12;
   e) a SOCS19 of SEQ ID NO: 14; or
   f) a WDS12 of SEQ ID NO: 16.

3. A fusion protein comprising the polypeptide of claim 1 or 2.

4. A binding compound which specifically binds to the polypeptide of claim 1 or 2.

5. The binding compound of claim 4 which is an antibody or antibody fragment.

6. A nucleic acid encoding the polypeptide of claim 1 or 2.

7. An expression vector comprising the nucleic acid of claim 6.

8. A host cell comprising the vector of claim 7.

9. A process for recombinatly producing a polypeptide comprising culturing the host cell of claim 8 under conditions in which the polypeptide is expressed.

SEQUENCE LISTING

SEQ ID NO: 1 is primate SOCS14 nucleic acid sequence.
SEQ ID NO: 2 is primate SOCS14 amino acid sequence.
SEQ ID NO: 3 is rodent SOCS15 (WDS11) nucleic acid sequence.
SEQ ID NO: 4 is rodent SOCS15 (WDS11) amino acid sequence.
SEQ ID NO: 5 is primate SOCS14 nucleic acid sequence.
SEQ ID NO: 6 is primate SOCS14 nucleic acid sequence.
SEQ ID NO: 7 is primate SOCS15 (WDS11) amino acid sequence.
SEQ ID NO: 8 is primate SOCS15 (WDS11) nucleic acid sequence.
SEQ ID NO: 9 is rodent SOCS17 amino acid sequence.
SEQ ID NO: 10 is rodent SOCS17 nucleic acid sequence.
SEQ ID NO: 11 is primate SOCS18 amino acid sequence.
SEQ ID NO: 12 is primate SOCS18 nucleic acid sequence.
SEQ ID NO: 13 is primate SOCS19 nucleic acid sequence.
SEQ ID NO: 14 is primate SOCS19 amino acid sequence.
SEQ ID NO: 15 is primate WDS12 nucleic acid sequence.
SEQ ID NO: 16 is mouse WDS12 amino acid sequence.
SEQ ID NO: 17 is mouse CIS amino acid sequence.
SEQ ID NO: 18 is primate SOCS1 amino acid sequence.
SEQ ID NO: 19 is murine SOCS1 amino acid sequence.
SEQ ID NO: 20 is primate SOCS2 amino acid sequence.
SEQ ID NO: 21 is primate SOCS3 amino acid sequence.
SEQ ID NO: 22 is murine SOCS3 amino acid sequence.
SEQ ID NO: 23 is primate SOCS16 amino acid sequence.
SEQ ID NO: 24 is primate SOCS14 nucleotide sequence.
SEQ ID NO: 25 is primate SOCS15 (WDS11) nucleotide sequence.
SEQ ID NO: 26 is rodent SOCS17 nucleotide sequence.
SEQ ID NO: 27 is primate SOCS18 nucleotide sequence.
SEQ ID NO: 28 is primate SOCS19 nucleotide sequence.
SEQ ID NO: 29 is primate WDS12 nucleotide sequence.

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Schering Corporation
(B) STREET: 2000 Galloping Hill Road
(C) CITY: Kenilworth
(D) STATE: New Jersey
(E) COUNTRY: USA
(F) POSTAL CODE: 07033-0530

(ii) TITLE OF INVENTION: Suppressors of Cytokine Signaling; Related Reagents

(iii) NUMBER OF SEQUENCES: 29

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: Apple Macintosh
(C) OPERATING SYSTEM: Macintosh 8.0.1
(D) SOFTWARE: Microsoft Word 6.0

(v) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE: 17-JUL-1998
(C) CLASSIFICATION:

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/055,804
(B) FILING DATE: 15-AUG-1997

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/055,853
(B) FILING DATE: 15-AUG-1997

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/053,153
(B) FILING DATE: 18-JUL-1997

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/053,244
(B) FILING DATE: 18-JUL-1997

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 930 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 824
(D) OTHER INFORMATION: /note= "position 824 is ambiguous; may be A, C, G, or T; all code for proline"

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..929

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
AC GAC CTC CAG TCT GAG ACC ACG TGC CAG GAG CAA GCC AAT TCA CTG        47
   Asp Leu Gln Ser Glu Thr Thr Cys Gln Glu Gln Ala Asn Ser Leu
     1               5                  10                  15

AAG AGC TCG GCT TCT CAT AAT GGA GAC CTG CAT CTT CAC CTG GAT GAA        95
Lys Ser Ser Ala Ser His Asn Gly Asp Leu His Leu His Leu Asp Glu
                 20                  25                  30

CAT GTG CCT GTC GTT ATT GGA CTT ATG CCT CAG GAC TAC ATT CAG TAT       143
His Val Pro Val Val Ile Gly Leu Met Pro Gln Asp Tyr Ile Gln Tyr
             35                  40                  45

ACT GTG CCT TTA GAT GAG GGG ATG TAT CCT TTG GAA GGA TCA CGG AGC       191
Thr Val Pro Leu Asp Glu Gly Met Tyr Pro Leu Glu Gly Ser Arg Ser
         50                  55                  60

TAT TGT CTG GAC AGC TCT TCT CCC ATG GAA GTC TCT GCG GTT CCT CCT       239
Tyr Cys Leu Asp Ser Ser Ser Pro Met Glu Val Ser Ala Val Pro Pro
     65                  70                  75

CAA GTG GGA GGG CGC GCT TTC CCC GAG GAT GAG AGT CAG GTA GAC CAG       287
Gln Val Gly Gly Arg Ala Phe Pro Glu Asp Glu Ser Gln Val Asp Gln
```

```
 80                  85                  90                  95

GAC CTA GTT GTC GCC CCA GAG ATC TTC GTG GAT CAG TCC GGT GAA TGG       335
Asp Leu Val Val Ala Pro Glu Ile Phe Val Asp Gln Ser Gly Glu Trp
                100                 105                 110

CTT GTT GAT TGG CAC CAC GGG AGT CAT GTT GCA GAA CCC CGG AGA GCG       383
Leu Val Asp Trp His His Gly Ser His Val Ala Glu Pro Arg Arg Ala
            115                 120                 125

GGT TCA CGA TGG ATG TCC CTC CAA TCT TCA CCA TTG GTT ACC TCC AAT       431
Gly Ser Arg Trp Met Ser Leu Gln Ser Ser Pro Leu Val Thr Ser Asn
        130                 135                 140

GCA GGA ATA ATC CAA ATC CCA AAG GGG ACC TTC AGT GGA CTC ACT GGG       479
Ala Gly Ile Ile Gln Ile Pro Lys Gly Thr Phe Ser Gly Leu Thr Gly
    145                 150                 155

ACA GAA GCC CAC GTG GCT GAA AGT ATG CGC TGT CAT TTG AAT TTT GAT       527
Thr Glu Ala His Val Ala Glu Ser Met Arg Cys His Leu Asn Phe Asp
160                 165                 170                 175

CCG AAC TCT GCT CCT GGG GTT GCA AGA GTT TAT GAC TCA GTG CAA AGT       575
Pro Asn Ser Ala Pro Gly Val Ala Arg Val Tyr Asp Ser Val Gln Ser
                180                 185                 190

AGT GGT CCC ATG GTT GTG ACA AGC CTT ACA GAG GAG CTG AAA AAA CTT       623
Ser Gly Pro Met Val Val Thr Ser Leu Thr Glu Glu Leu Lys Lys Leu
            195                 200                 205

GCA AAG CAA GGA TGG TAC TGG GGA CCA ATC ACA CGT TGG GAG GCA GAA       671
Ala Lys Gln Gly Trp Tyr Trp Gly Pro Ile Thr Arg Trp Glu Ala Glu
        210                 215                 220

GGG AAG CTA GCA AAC GTG CCA GAT GGT TCT TTT CTT GTT CGG GAC AGT       719
Gly Lys Leu Ala Asn Val Pro Asp Gly Ser Phe Leu Val Arg Asp Ser
    225                 230                 235

TCT GAC GAC CGT TAC CTT TTA AGC TTG AGC TTT CGC TCC CAT GGT AAA       767
Ser Asp Asp Arg Tyr Leu Leu Ser Leu Ser Phe Arg Ser His Gly Lys
240                 245                 250                 255

ACA CTT CAC ACT AGA ATT GAG CAC TCA AAT GGT AGG TTT AGC TTT TAT       815
Thr Leu His Thr Arg Ile Glu His Ser Asn Gly Arg Phe Ser Phe Tyr
                260                 265                 270

GAA CAG CCC GAT GTG GAA GGA CAT ACG TCC ATA GTT GAT CTA ATT GGA       863
Glu Gln Pro Asp Val Glu Gly His Thr Ser Ile Val Asp Leu Ile Gly
            275                 280                 285

GCA TTC AAT CAG GGA CTC TGA AAA TGG GAG CTT TTT GTT ATT CAA GGT       911
Ala Phe Asn Gln Gly Leu  *  Lys Trp Glu Leu Phe Val Ile Gln Gly
        290                 295                 300

CTC GGC TGC CTG GAA TCT G                                             930
Leu Gly Cys Leu Glu Ser
    305
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 309 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Asp Leu Gln Ser Glu Thr Thr Cys Gln Glu Gln Ala Asn Ser Leu Lys
 1               5                  10                  15

Ser Ser Ala Ser His Asn Gly Asp Leu His Leu His Leu Asp Glu His
            20                  25                  30

Val Pro Val Val Ile Gly Leu Met Pro Gln Asp Tyr Ile Gln Tyr Thr
        35                  40                  45

Val Pro Leu Asp Glu Gly Met Tyr Pro Leu Glu Gly Ser Arg Ser Tyr
    50                  55                  60

Cys Leu Asp Ser Ser Ser Pro Met Glu Val Ser Ala Val Pro Pro Gln
65                  70                  75                  80

Val Gly Gly Arg Ala Phe Pro Glu Asp Glu Ser Gln Val Asp Gln Asp
                85                  90                  95

Leu Val Val Ala Pro Glu Ile Phe Val Asp Gln Ser Gly Glu Trp Leu
            100                 105                 110

Val Asp Trp His His Gly Ser His Val Ala Glu Pro Arg Arg Ala Gly
        115                 120                 125

Ser Arg Trp Met Ser Leu Gln Ser Ser Pro Leu Val Thr Ser Asn Ala
    130                 135                 140

Gly Ile Ile Gln Ile Pro Lys Gly Thr Phe Ser Gly Leu Thr Gly Thr
145                 150                 155                 160

Glu Ala His Val Ala Glu Ser Met Arg Cys His Leu Asn Phe Asp Pro
                165                 170                 175

Asn Ser Ala Pro Gly Val Ala Arg Val Tyr Asp Ser Val Gln Ser Ser
            180                 185                 190

Gly Pro Met Val Val Thr Ser Leu Thr Glu Glu Leu Lys Lys Leu Ala
        195                 200                 205

Lys Gln Gly Trp Tyr Trp Gly Pro Ile Thr Arg Trp Glu Ala Glu Gly
    210                 215                 220

Lys Leu Ala Asn Val Pro Asp Gly Ser Phe Leu Val Arg Asp Ser Ser
225                 230                 235                 240

Asp Asp Arg Tyr Leu Leu Ser Leu Ser Phe Arg Ser His Gly Lys Thr
                245                 250                 255

Leu His Thr Arg Ile Glu His Ser Asn Gly Arg Phe Ser Phe Tyr Glu
            260                 265                 270

Gln Pro Asp Val Glu Gly His Thr Ser Ile Val Asp Leu Ile Gly Ala
        275                 280                 285

Phe Asn Gln Gly Leu  *  Lys Trp Glu Leu Phe Val Ile Gln Gly Leu
    290                 295                 300

Gly Cys Leu Glu Ser
```

305

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 476 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..476

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CA GCT TCG TAT GAC ACC AGT GTG ATT ATG TGG GAC CCC TAC ACC GGC        47
   Ala Ser Tyr Asp Thr Ser Val Ile Met Trp Asp Pro Tyr Thr Gly
     1               5                  10                  15

GAG AGG CTG AGG TCA CTT CAT CAC ACA CAG CTT GAA CCC ACC ATG GAT       95
Glu Arg Leu Arg Ser Leu His His Thr Gln Leu Glu Pro Thr Met Asp
                 20                  25                  30

GAC AGT GAC GTC CAC ATG AGC TCC CTG AGG TCC GTG TGC TTC TCA CCT      143
Asp Ser Asp Val His Met Ser Ser Leu Arg Ser Val Cys Phe Ser Pro
             35                  40                  45

GAA GGC TTG TAT CTC GCT ACG GTG GCA GAT GAC AGG CTG CTC AGG ATC      191
Glu Gly Leu Tyr Leu Ala Thr Val Ala Asp Asp Arg Leu Leu Arg Ile
         50                  55                  60

TGG GCT CTG GAA CTG AAG GCT CCG GTT GCC TTT GCT CCG ATG ACC AAT      239
Trp Ala Leu Glu Leu Lys Ala Pro Val Ala Phe Ala Pro Met Thr Asn
     65                  70                  75

GGT CTT TGC TGC ACG TTC TTC CCA CAC GGT GGA ATT ATT GCC ACA GGG      287
Gly Leu Cys Cys Thr Phe Phe Pro His Gly Gly Ile Ile Ala Thr Gly
 80                  85                  90                  95

ACG AGA GAT GGC CAT GTC CAG TTC TGG ACA GCT CCC CGG GTC CTG TCC      335
Thr Arg Asp Gly His Val Gln Phe Trp Thr Ala Pro Arg Val Leu Ser
                100                 105                 110

TCA CTG AAG CAC TTA TGC AGG AAA GCC CTC CGA AGT TTC CTG ACA ACG      383
Ser Leu Lys His Leu Cys Arg Lys Ala Leu Arg Ser Phe Leu Thr Thr
            115                 120                 125

TAT CAA GTC CTA GCA CTG CCA ATC CCC AAG AAG ATG AAA GAG TTC CTC      431
Tyr Gln Val Leu Ala Leu Pro Ile Pro Lys Lys Met Lys Glu Phe Leu
        130                 135                 140

ACA TAC AGG ACT TTC TAG CAG TGC CGG CTC CCC CAC CTC CTG CAG          476
Thr Tyr Arg Thr Phe  *  Gln Cys Arg Leu Pro His Leu Leu Gln
    145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 158 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Ala Ser Tyr Asp Thr Ser Val Ile Met Trp Asp Pro Tyr Thr Gly Glu
 1               5                  10                  15

Arg Leu Arg Ser Leu His His Thr Gln Leu Glu Pro Thr Met Asp Asp
            20                  25                  30

Ser Asp Val His Met Ser Ser Leu Arg Ser Val Cys Phe Ser Pro Glu
        35                  40                  45

Gly Leu Tyr Leu Ala Thr Val Ala Asp Asp Arg Leu Leu Arg Ile Trp
    50                  55                  60

Ala Leu Glu Leu Lys Ala Pro Val Ala Phe Ala Pro Met Thr Asn Gly
65                  70                  75                  80

Leu Cys Cys Thr Phe Phe Pro His Gly Gly Ile Ile Ala Thr Gly Thr
                85                  90                  95

Arg Asp Gly His Val Gln Phe Trp Thr Ala Pro Arg Val Leu Ser Ser
            100                 105                 110

Leu Lys His Leu Cys Arg Lys Ala Leu Arg Ser Phe Leu Thr Thr Tyr
        115                 120                 125

Gln Val Leu Ala Leu Pro Ile Pro Lys Lys Met Lys Glu Phe Leu Thr
    130                 135                 140

Tyr Arg Thr Phe  *  Gln Cys Arg Leu Pro His Leu Leu Gln
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2093 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 87..1241

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 20
(D) OTHER INFORMATION: /note= "nucleotide may be A or C at positions: 20, 36, 1583, 1675, 1689, 1693, 1710, 1711, 1719, 1720, 1728, 1753, 1787, and 1806."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 35
(D) OTHER INFORMATION: /note= "nucleotide may be G or T at

postions: 35, 1541, 1594, 1689, 1778, 1779, 1825, 1844, 1845, 1853, 1854, 1865, 1884, and 1893."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 70
(D) OTHER INFORMATION: /note= "Nucleotide may be A or G at positions: 70, 1461, 1630, 1677, 1713, 1725, 1734, 1735, 1757, 1805, 1810, and 1863."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 64
(D) OTHER INFORMATION: /note= "Nucleotide may be A or T at positions: 64, 1692, 1715, 1718, 1721, 1722, 1799, 1837, 1841, 1876, and 1894."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1661
(D) OTHER INFORMATION: /note= "Nucleotide may be C or T at positions: 1661, 1729, 1749, 1750, 1754, 1776, 1802, 1826, 1847, 1859, 1860, 1904, 1907, and 1911."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1731
(D) OTHER INFORMATION: /note= "Nucleotide may be G or C at positions: 1731, 1817, 1887, and 1908."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1869
(D) OTHER INFORMATION: /note= "Nucleotide may be C, G, or T at positions: 1869, 1883, 1885, 1886, and 1895."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1888
(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, or G at positions: 1888, and 1896."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1877
(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, or T at positions: 1877, and 1898."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1855
(D) OTHER INFORMATION: /note= "Nucleotide may be A, G, or T at position 1855."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1935
(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, G, or T at positions: 1935, and 2034."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TAAGGTCCAC GTCGCTCCGC AGCCATCACT ACAGGCCCGC GCCGTGGCCT CTGCGGCCCA 60

```
CAATCTCCGA GGAGACCTGC ATCAAG ATG GAG GTG AGA GTC AAG GCC TTG GTT     113
                             Met Glu Val Arg Val Lys Ala Leu Val
                               1               5

CAC TCT TCC AGC CCG AGT CCA GCC CTG AAT GGC GTC CGG AAG GAT TTC      161
His Ser Ser Ser Pro Ser Pro Ala Leu Asn Gly Val Arg Lys Asp Phe
 10                  15                  20                  25

CAC GAC CTC CAG TCT GAG ACC ACG TGC CAG GAG CAA GCC AAT TCA CTG      209
His Asp Leu Gln Ser Glu Thr Thr Cys Gln Glu Gln Ala Asn Ser Leu
                 30                  35                  40

AAG AGC TCG GCT TCT CAT AAT GGA GAC CTG CAT CTT CAC CTG GAT GAA      257
Lys Ser Ser Ala Ser His Asn Gly Asp Leu His Leu His Leu Asp Glu
             45                  50                  55

CAT GTG CCT GTC GTT ATT GGA CTT ATG CCT CAG GAC TAC ATT CAG TAT      305
His Val Pro Val Val Ile Gly Leu Met Pro Gln Asp Tyr Ile Gln Tyr
         60                  65                  70

ACT GTG CCT TTA GAT GAG GGG ATG TAT CCT TTG GAA GGA TCA CGG AGC      353
Thr Val Pro Leu Asp Glu Gly Met Tyr Pro Leu Glu Gly Ser Arg Ser
     75                  80                  85

TAT TGT CTG GAC AGC TCT TCT CCC ATG GAA GTC TCT GCG GTT CCT CCT      401
Tyr Cys Leu Asp Ser Ser Ser Pro Met Glu Val Ser Ala Val Pro Pro
 90                  95                 100                 105

CAA GTG GGA GGG CGC GCT TTC CCC GAG GAT GAG AGT CAG GTA GAC CAG      449
Gln Val Gly Gly Arg Ala Phe Pro Glu Asp Glu Ser Gln Val Asp Gln
                110                 115                 120

GAC CTA GTT GTC GCC CCA GAG ATC TTC GTG GAT CAG TCC GTG AAT GGC      497
Asp Leu Val Val Ala Pro Glu Ile Phe Val Asp Gln Ser Val Asn Gly
            125                 130                 135

TTG TTG ATT GGC ACC ACG GGA GTC ATG TTG CAG AGC CCG AGA GCG GGT      545
Leu Leu Ile Gly Thr Thr Gly Val Met Leu Gln Ser Pro Arg Ala Gly
        140                 145                 150

CAC GAT GAT GTC CCT CCA CTC TCA CCA TTG CTA CCT CCA ATG CAG AAT      593
His Asp Asp Val Pro Pro Leu Ser Pro Leu Leu Pro Pro Met Gln Asn
    155                 160                 165

AAT CAA ATC CAA AGG AAC TTC AGT GGA CTC ACT GGC ACA GAA GCC CAC      641
Asn Gln Ile Gln Arg Asn Phe Ser Gly Leu Thr Gly Thr Glu Ala His
170                 175                 180                 185

GTG GCT GAA AGT ATG CGC TGT CAT TTG AAT TTT GAT CCG AAC TCT GCT      689
Val Ala Glu Ser Met Arg Cys His Leu Asn Phe Asp Pro Asn Ser Ala
                190                 195                 200

CCT GGG GTT GCA AGA GTT TAT GAC TCA GTG CAA AGT AGT GGT CCC ATG      737
Pro Gly Val Ala Arg Val Tyr Asp Ser Val Gln Ser Ser Gly Pro Met
            205                 210                 215

GTT GTG ACA AGC CTT ACA GAG GAG CTG AAA AAA CTT GCA AAG CAA GGA      785
Val Val Thr Ser Leu Thr Glu Glu Leu Lys Lys Leu Ala Lys Gln Gly
        220                 225                 230

TGG TAC TGG GGA CCA ATC ACA CGT TGG GAG GCA GAA GGG AAG CTA GCA      833
Trp Tyr Trp Gly Pro Ile Thr Arg Trp Glu Ala Glu Gly Lys Leu Ala
    235                 240                 245
```

```
AAC GTG CCA GAT GGT TCT TTT CTT GTT CGG GAC AGT TCT GAC GAC CGT        881
Asn Val Pro Asp Gly Ser Phe Leu Val Arg Asp Ser Ser Asp Asp Arg
250                 255                 260                 265

TAC CTT TTA AGC TTG AGC TTT CGC TCC CAT GGT AAA ACA CTT CAC ACT        929
Tyr Leu Leu Ser Leu Ser Phe Arg Ser His Gly Lys Thr Leu His Thr
                270                 275                 280

AGA ATT GAG CAC TCA AAT GGT AGG TTT AGC TTT TAT GAA CAG CCA GAT        977
Arg Ile Glu His Ser Asn Gly Arg Phe Ser Phe Tyr Glu Gln Pro Asp
            285                 290                 295

GTG GAA AGG ACA TAC TCC ATA GTT GAT CTA ATT GAG CAT TCC ATC CAG       1025
Val Glu Arg Thr Tyr Ser Ile Val Asp Leu Ile Glu His Ser Ile Gln
        300                 305                 310

GGA CTC GAA AAT GGA GCT TTT TGT TAT TCA AGG TCT CGG CTG CCT GGA       1073
Gly Leu Glu Asn Gly Ala Phe Cys Tyr Ser Arg Ser Arg Leu Pro Gly
    315                 320                 325

TCT GCA ACT TAC CCC GTC AGA CTG ACC AAC CCA GTG TCC CGG TTC ATG       1121
Ser Ala Thr Tyr Pro Val Arg Leu Thr Asn Pro Val Ser Arg Phe Met
330                 335                 340                 345

CAG GTG CGC TCG TTG CAG TAC CTG TGT CGT TTT GTT ATA CGT CAG TAT       1169
Gln Val Arg Ser Leu Gln Tyr Leu Cys Arg Phe Val Ile Arg Gln Tyr
                350                 355                 360

ACC AGA ATA GAC TTA ATT CAG AAA CTG CCT TTG CCA AAC AAA ATG AAG       1217
Thr Arg Ile Asp Leu Ile Gln Lys Leu Pro Leu Pro Asn Lys Met Lys
            365                 370                 375

GAT TAT TTA CAG GAG AAG CAC TAC TGAAAGATTG AGAACCCTGC ATCTTGCACT      1271
Asp Tyr Leu Gln Glu Lys His Tyr
        380                 385

TTGGGAATAA GAACAAGAGA TTGAAATACA GTTTACAAAC TTTCATTGCC ATCAAAATCT     1331

TTTGCTGCCA TAACTATTTC AGTTTTATGT GTAAAAGAGT CATCAGTTTG TTTAGGGGTG     1391

GGGAAGTGTC AGCAAGGTGT CTTGGGTTTA TTTTGGTTCT TTAAAAAAGG GAAGTCTTGA     1451

AGTTTTAGAA GTGTTGAATT ATGTTTCATC AATGTGCAGA ATAATCACAA TGTGAATTAT     1511

CAAATTCTCC TCAATGCCCC CCCCGCCCAT TCCTTTGCTG CTATCCACTG TGATTTTTAT     1571

GCATTAAAAG CCCATTTCAT GTTTTTTCAA CCCTAAGTAA AGTTGAATGA AACTTAACAG     1631

AATGGAAATT GCTATTTCTT TTTAAATGGC CCATTTTCCA AAACAAGTGT TGAATAACCA     1691

ACCCTGTTTG AATAAAACCC GAAATTACCA ATAACACCGG AGGTGAGTTT TTAATCTCCT     1751

ACCTTGAAAA GATTTATTTA GAATCGGGAA TTGACCTAAT ATTGGGTAAT TGGACCGGAG     1811

ATCTGCAACA TATTCTTTAA CAACAATTTA TTGGCCTTAA TTTGTTTCCA AAGGTGGCCT     1871

TATTTCTTTG GGGGGGGAAA GGAGGAATTC TCCGTCCCCC TCGTTTTCAT CTTCTAGTTT     1931

GTGCTATTTT AATAAATGGC CTTACATTAA AAAATTGTAA AGAAATGTAT ACCACCAATT     1991

TAGAAATTGT TGCCTTTTCT GTAATTAAAC TCGGGTACAA ATCGGCATAA CATGAAAACC     2051

TATGGAACTA GAATTATTAT TAAAGAAATA TTAGATGATC AT                        2093
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 385 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Glu Val Arg Val Lys Ala Leu Val His Ser Ser Ser Pro Ser Pro
1               5                   10                  15

Ala Leu Asn Gly Val Arg Lys Asp Phe His Asp Leu Gln Ser Glu Thr
            20                  25                  30

Thr Cys Gln Glu Gln Ala Asn Ser Leu Lys Ser Ser Ala Ser His Asn
        35                  40                  45

Gly Asp Leu His Leu His Leu Asp Glu His Val Pro Val Val Ile Gly
    50                  55                  60

Leu Met Pro Gln Asp Tyr Ile Gln Tyr Thr Val Pro Leu Asp Glu Gly
65                  70                  75                  80

Met Tyr Pro Leu Glu Gly Ser Arg Ser Tyr Cys Leu Asp Ser Ser Ser
                85                  90                  95

Pro Met Glu Val Ser Ala Val Pro Pro Gln Val Gly Gly Arg Ala Phe
            100                 105                 110

Pro Glu Asp Glu Ser Gln Val Asp Gln Asp Leu Val Val Ala Pro Glu
        115                 120                 125

Ile Phe Val Asp Gln Ser Val Asn Gly Leu Leu Ile Gly Thr Thr Gly
    130                 135                 140

Val Met Leu Gln Ser Pro Arg Ala Gly His Asp Asp Val Pro Pro Leu
145                 150                 155                 160

Ser Pro Leu Leu Pro Pro Met Gln Asn Asn Gln Ile Gln Arg Asn Phe
                165                 170                 175

Ser Gly Leu Thr Gly Thr Glu Ala His Val Ala Glu Ser Met Arg Cys
            180                 185                 190

His Leu Asn Phe Asp Pro Asn Ser Ala Pro Gly Val Ala Arg Val Tyr
        195                 200                 205

Asp Ser Val Gln Ser Ser Gly Pro Met Val Val Thr Ser Leu Thr Glu
    210                 215                 220

Glu Leu Lys Lys Leu Ala Lys Gln Gly Trp Tyr Trp Gly Pro Ile Thr
225                 230                 235                 240

Arg Trp Glu Ala Glu Gly Lys Leu Ala Asn Val Pro Asp Gly Ser Phe
                245                 250                 255

Leu Val Arg Asp Ser Ser Asp Asp Arg Tyr Leu Leu Ser Leu Ser Phe
            260                 265                 270
```

```
Arg Ser His Gly Lys Thr Leu His Thr Arg Ile Glu His Ser Asn Gly
        275                 280                 285

Arg Phe Ser Phe Tyr Glu Gln Pro Asp Val Glu Arg Thr Tyr Ser Ile
    290                 295                 300

Val Asp Leu Ile Glu His Ser Ile Gln Gly Leu Glu Asn Gly Ala Phe
305                 310                 315                 320

Cys Tyr Ser Arg Ser Arg Leu Pro Gly Ser Ala Thr Tyr Pro Val Arg
                325                 330                 335

Leu Thr Asn Pro Val Ser Arg Phe Met Gln Val Arg Ser Leu Gln Tyr
            340                 345                 350

Leu Cys Arg Phe Val Ile Arg Gln Tyr Thr Arg Ile Asp Leu Ile Gln
        355                 360                 365

Lys Leu Pro Leu Pro Asn Lys Met Lys Asp Tyr Leu Gln Glu Lys His
    370                 375                 380

Tyr
385
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1748 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1335

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1026
(D) OTHER INFORMATION: /note= "Nucleotide may be C or T at positions: 1026, 1032, 1041, 1452, 1510, and 1567."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 945
(D) OTHER INFORMATION: /note= "Nucleotide may be A or G at positions: 945, 1376, 1541, 1658, 1662, and 1668."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1435
(D) OTHER INFORMATION: /note= "Nucleotide may be G or T at positions: 1435, 1481, 1518, and 1543."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1500
(D) OTHER INFORMATION: /note= "Nucleotide may be A or C at positions: 1500, and 1669."

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION: 1521
(D) OTHER INFORMATION: /note= "Nucleotide may be A or T at positions: 1521, and 1542."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1651
(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, or T at position 1651."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1654
(D) OTHER INFORMATION: /note= "Nucleotide may be G, T, or C at position 1654."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1656
(D) OTHER INFORMATION: /note= "Nucleotide may be G, C, or A at position 1656."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1589..1649
(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, T, or G at positions: 1589-1649, 1652, 1655, 1657-1661, 1664-1667, and 1672-1748."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
ATG GAG GCC GGA GAG GAA CCG CTG CTG CTG GCC GAA CTC AAG CCC GGG       48
Met Glu Ala Gly Glu Glu Pro Leu Leu Leu Ala Glu Leu Lys Pro Gly
  1               5                  10                  15

CGC CCC CAC CAG TTT GAT TGG AAG TCC AGC TGT GAA ACC TGG AGC GTG       96
Arg Pro His Gln Phe Asp Trp Lys Ser Ser Cys Glu Thr Trp Ser Val
             20                  25                  30

GCC TTC TCG CCA GAC GGT TCC TGG TTC GCC TGG TCT CAA GGA CAC TGC      144
Ala Phe Ser Pro Asp Gly Ser Trp Phe Ala Trp Ser Gln Gly His Cys
         35                  40                  45

GTG GTC AAG CTG GTC CCC TGG CCC TTA GAG GAA CAG TTC ATC CCT AAA      192
Val Val Lys Leu Val Pro Trp Pro Leu Glu Glu Gln Phe Ile Pro Lys
     50                  55                  60

GGA TTC GAA GCC AAG AGC CGA AGC AGC AAG AAT GAC CCA AAA GGA CGG      240
Gly Phe Glu Ala Lys Ser Arg Ser Ser Lys Asn Asp Pro Lys Gly Arg
 65                  70                  75                  80

GGC AGT CTG AAG GAG AAG ACG CTG GAC TGT GGC CAG ATT GTG TGG GGG      288
Gly Ser Leu Lys Glu Lys Thr Leu Asp Cys Gly Gln Ile Val Trp Gly
                 85                  90                  95

CTG GCC TTC AGC CCA TGG CCC TCT CCA CCC AGC AGG AAA CTC TGG GCA      336
Leu Ala Phe Ser Pro Trp Pro Ser Pro Pro Ser Arg Lys Leu Trp Ala
            100                 105                 110

CGT CAC CAT CCC CAG GCG CCT GAT GTT TCT TGC CTG ATC CTG GCC ACA      384
Arg His His Pro Gln Ala Pro Asp Val Ser Cys Leu Ile Leu Ala Thr
        115                 120                 125
```

```
GGT CTC AAC GAT GGG CAG ATC AAG ATT TGG GAG GTA CAG ACA GGC CTC        432
Gly Leu Asn Asp Gly Gln Ile Lys Ile Trp Glu Val Gln Thr Gly Leu
    130                 135                     140

CTG CTT CTG AAT CTT TCT GGC CAC CAA GAC GTC GTG AGA GAT CTG AGC        480
Leu Leu Leu Asn Leu Ser Gly His Gln Asp Val Val Arg Asp Leu Ser
145                 150                     155                 160

TTC ACG CCC AGC GGC AGT TTG ATT TTG GTC TCT GCA TCC CGG GAT AAG        528
Phe Thr Pro Ser Gly Ser Leu Ile Leu Val Ser Ala Ser Arg Asp Lys
                165                     170                 175

ACA CTT CGA ATT TGG GAC CTG AAT AAG CAC GGT AAG CAG ATC CAG GTG        576
Thr Leu Arg Ile Trp Asp Leu Asn Lys His Gly Lys Gln Ile Gln Val
            180                 185                     190

TTA TCC GGC CAT CTG CAG TGG GTT TAC TGC TGC TCC ATC TCC CCT GAC        624
Leu Ser Gly His Leu Gln Trp Val Tyr Cys Cys Ser Ile Ser Pro Asp
        195                 200                     205

TGT AGC ATG CTG TGC TCT GCA GCT GGG GAG AAG TCG GTC TTT CTG TGG        672
Cys Ser Met Leu Cys Ser Ala Ala Gly Glu Lys Ser Val Phe Leu Trp
    210                 215                     220

AGC ATG CGG TCC TAC ACA CTA ATC CGG AAA CTA GAA GGC CAC CAA AGC        720
Ser Met Arg Ser Tyr Thr Leu Ile Arg Lys Leu Glu Gly His Gln Ser
225                 230                     235                 240

AGT GTT GTC TCC TGT GAT TTC TCT CCT GAT TCA GCC TTG CTT GTC ACA        768
Ser Val Val Ser Cys Asp Phe Ser Pro Asp Ser Ala Leu Leu Val Thr
                245                     250                 255

GCT TCG TAT GAC ACC AGT GTG ATT ATG TGG GAC CCC TAC ACC GGC GAG        816
Ala Ser Tyr Asp Thr Ser Val Ile Met Trp Asp Pro Tyr Thr Gly Glu
            260                 265                     270

AGG CTG AGG TCA CTT CAT CAC ACA CAG CTT GAA CCC ACC ATG GAT GAC        864
Arg Leu Arg Ser Leu His His Thr Gln Leu Glu Pro Thr Met Asp Asp
        275                 280                     285

AGT GAC GTC CAC ATG AGC TCC CTG AGG TCC GTG TGC TTC TCA CCT GAA        912
Ser Asp Val His Met Ser Ser Leu Arg Ser Val Cys Phe Ser Pro Glu
    290                 295                     300

GGC TTG TAT CTC GCT ACG GTG GCA GAT GAC AGA CTG CTC AGG ATC TGG        960
Gly Leu Tyr Leu Ala Thr Val Ala Asp Asp Arg Leu Leu Arg Ile Trp
305                 310                     315                 320

GCT CTG GAA CTG AAA GCT CCG GTT GCC TTT GCT CCG ATG ACC AAT GGT       1008
Ala Leu Glu Leu Lys Ala Pro Val Ala Phe Ala Pro Met Thr Asn Gly
                325                     330                 335

CTT TGC TGC ACA TTT TTC CCA CAC GGT GGA ATC ATT GCC ACA GGG ACA       1056
Leu Cys Cys Thr Phe Phe Pro His Gly Gly Ile Ile Ala Thr Gly Thr
            340                 345                     350

AGA GAT GGC CAC GTC CAG TTC TGG ACA GCT CCT AGG GTC CTG TCC TCA       1104
Arg Asp Gly His Val Gln Phe Trp Thr Ala Pro Arg Val Leu Ser Ser
        355                 360                     365

CTG AAG CAC TTA TGC CGG AAA GCC CTT CGA AGT TTC CTA ACA ACT TAC       1152
Leu Lys His Leu Cys Arg Lys Ala Leu Arg Ser Phe Leu Thr Thr Tyr
    370                 375                     380
```

```
CAA GTC CTA GCA CTG CCA ATC CCC AAG AAA ATG AAA GAG TTC CTC ACA     1200
Gln Val Leu Ala Leu Pro Ile Pro Lys Lys Met Lys Glu Phe Leu Thr
385                 390                 395                 400

TAC AGG ACT TTT TAA GCA ACA CCA CAT CTT GTG CTT CTT TGT AGC AGG     1248
Tyr Arg Thr Phe  *  Ala Thr Pro His Leu Val Leu Leu Cys Ser Arg
                405                 410                 415

GTA AAT CGT CCT GTC AAA GGG AGT TGC TGG AAT AAT GGG CCA AAC ATC     1296
Val Asn Arg Pro Val Lys Gly Ser Cys Trp Asn Asn Gly Pro Asn Ile
            420                 425                 430

TGG TCT TGC ATT GAA ATA GCA TTT CTT TGG GAT TGT GAA TAGAATGTAG      1345
Trp Ser Cys Ile Glu Ile Ala Phe Leu Trp Asp Cys Glu
        435                 440                 445

CAAAACCAGA TTCCAGTGTA CTAGTCATGG GTCTTTCTCT CCCTGGGCAT GTGGAAAGTC   1405

AGTCTTAGGA GGGAAGGAGA TTCCACTTGG CACGGGCAAC AGAGCCCTTA CGTTTAAATT   1465

TTTCAGTCCA GTTATTGAAC AGCAAGTGTT TGAAATCTTT CTGGCTTGTT TTGGATTTCA   1525

AAGTGGCAGT TACTGGTGGT TGTTTTTGGA TTTATGGCAA CCAAGTTAGG GCCTCCAGCG   1585

GTTCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC   1645

CCCCTCCACC CCGCCCATCC CCACATCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC   1705

CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCC                     1748
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 445 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Glu Ala Gly Glu Glu Pro Leu Leu Leu Ala Glu Leu Lys Pro Gly
 1               5                  10                  15

Arg Pro His Gln Phe Asp Trp Lys Ser Ser Cys Glu Thr Trp Ser Val
            20                  25                  30

Ala Phe Ser Pro Asp Gly Ser Trp Phe Ala Trp Ser Gln Gly His Cys
        35                  40                  45

Val Val Lys Leu Val Pro Trp Pro Leu Glu Glu Gln Phe Ile Pro Lys
    50                  55                  60

Gly Phe Glu Ala Lys Ser Arg Ser Ser Lys Asn Asp Pro Lys Gly Arg
65                  70                  75                  80

Gly Ser Leu Lys Glu Lys Thr Leu Asp Cys Gly Gln Ile Val Trp Gly
                85                  90                  95

Leu Ala Phe Ser Pro Trp Pro Ser Pro Pro Ser Arg Lys Leu Trp Ala
            100                 105                 110
```

```
Arg His His Pro Gln Ala Pro Asp Val Ser Cys Leu Ile Leu Ala Thr
        115                 120                 125

Gly Leu Asn Asp Gly Gln Ile Lys Ile Trp Glu Val Gln Thr Gly Leu
    130                 135                 140

Leu Leu Leu Asn Leu Ser Gly His Gln Asp Val Val Arg Asp Leu Ser
145                 150                 155                 160

Phe Thr Pro Ser Gly Ser Leu Ile Leu Val Ser Ala Ser Arg Asp Lys
                165                 170                 175

Thr Leu Arg Ile Trp Asp Leu Asn Lys His Gly Lys Gln Ile Gln Val
            180                 185                 190

Leu Ser Gly His Leu Gln Trp Val Tyr Cys Cys Ser Ile Ser Pro Asp
        195                 200                 205

Cys Ser Met Leu Cys Ser Ala Ala Gly Glu Lys Ser Val Phe Leu Trp
    210                 215                 220

Ser Met Arg Ser Tyr Thr Leu Ile Arg Lys Leu Glu Gly His Gln Ser
225                 230                 235                 240

Ser Val Val Ser Cys Asp Phe Ser Pro Asp Ser Ala Leu Leu Val Thr
                245                 250                 255

Ala Ser Tyr Asp Thr Ser Val Ile Met Trp Asp Pro Tyr Thr Gly Glu
            260                 265                 270

Arg Leu Arg Ser Leu His His Thr Gln Leu Glu Pro Thr Met Asp Asp
        275                 280                 285

Ser Asp Val His Met Ser Ser Leu Arg Ser Val Cys Phe Ser Pro Glu
    290                 295                 300

Gly Leu Tyr Leu Ala Thr Val Ala Asp Asp Arg Leu Leu Arg Ile Trp
305                 310                 315                 320

Ala Leu Glu Leu Lys Ala Pro Val Ala Phe Ala Pro Met Thr Asn Gly
                325                 330                 335

Leu Cys Cys Thr Phe Phe Pro His Gly Gly Ile Ile Ala Thr Gly Thr
            340                 345                 350

Arg Asp Gly His Val Gln Phe Trp Thr Ala Pro Arg Val Leu Ser Ser
        355                 360                 365

Leu Lys His Leu Cys Arg Lys Ala Leu Arg Ser Phe Leu Thr Thr Tyr
    370                 375                 380

Gln Val Leu Ala Leu Pro Ile Pro Lys Lys Met Lys Glu Phe Leu Thr
385                 390                 395                 400

Tyr Arg Thr Phe  *  Ala Thr Pro His Leu Val Leu Leu Cys Ser Arg
                405                 410                 415

Val Asn Arg Pro Val Lys Gly Ser Cys Trp Asn Asn Gly Pro Asn Ile
            420                 425                 430

Trp Ser Cys Ile Glu Ile Ala Phe Leu Trp Asp Cys Glu
        435                 440                 445
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2198 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1419

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1680
(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, T, or G at positions: 1680, 1691, 1696, 1704, 1707, 1728, 1740, 1743, 1746, 1755, 1760, 1770, 1773, 1802, 1816, 1817, 1823, 1826, 1827, 1846, 1851, 1857, 1861, 1880, and 1885."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1909
(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, T, or G at positions: 1909, 1917, 1920, 1929, 1946, 1953, 1967-8, 1980, 1991, 1995, 2001, 2004, 2021, 2033-37, 2039-40, 2042, 2048, 2051, 2054, 2061, 2075, 2081, and 2083-85. "

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 2088
(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, T, or G at positions: 2088, 2105, 2121, 2124, 2132, 2137, 2147, 2149, 2151-52, 2160, 2165, 2177, 2179 and 2196."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 494
(D) OTHER INFORMATION: /note= "Nucleotide may be A or C at position 494."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 498
(D) OTHER INFORMATION: /note= "Nucleotide may be C or T at positions: 498, 501, 1455, 1524, 1527, 1621, 1829, and 2072."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 499
(D) OTHER INFORMATION: /note= "Nucleotide may be G or C at positions: 499, 1618, and 1664."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1673
(D) OTHER INFORMATION: /note= "Nucleotide may be G or T at position 1673."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1819

(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, or G at positions: 1819, 1840, and 2089."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
GGC GGT GGT GAT GGC GGC AGG CGC TCG GAC AGC TCC GCT TGA GCT GAG        48
Gly Gly Gly Asp Gly Gly Arg Arg Ser Asp Ser Ser Ala  *  Ala Glu
 1               5                  10                  15

CTC GGA GAG ATC CGT CCA GAA AGT GCC CAG AAG AAA CTT CCT CTT AGA        96
Leu Gly Glu Ile Arg Pro Glu Ser Ala Gln Lys Lys Leu Pro Leu Arg
            20                  25                  30

AAA GCT GAA AAC ACA ATA TTT ATA ACA CTG GAA ATT GTA AAG AAT TTG       144
Lys Ala Glu Asn Thr Ile Phe Ile Thr Leu Glu Ile Val Lys Asn Leu
        35                  40                  45

TTT AAA ATG GCT GAA AAC AAT AGT AAA AAT GTA GAT GTA CGG CCT AAA       192
Phe Lys Met Ala Glu Asn Asn Ser Lys Asn Val Asp Val Arg Pro Lys
    50                  55                  60

ACA AGT CGG AGT CGA AGT GCT GAC AGG AAG GAT GGT TAT GTG TGG AGT       240
Thr Ser Arg Ser Arg Ser Ala Asp Arg Lys Asp Gly Tyr Val Trp Ser
65                  70                  75                  80

GGA AAG AAG TTG TCT TGG TCC AAA AAG AGT GAG AGT TGT TCT GAA TCT       288
Gly Lys Lys Leu Ser Trp Ser Lys Lys Ser Glu Ser Cys Ser Glu Ser
                85                  90                  95

GAA GCC AAG AAA GGG CAG CTT AGC TGT TCG TCC ATT GAG TTG GAC TTA       336
Glu Ala Lys Lys Gly Gln Leu Ser Cys Ser Ser Ile Glu Leu Asp Leu
            100                 105                 110

GAT CAT TCC TGT GGG CAT AGA TTT TTA GGC CGA TCC CTT AAA CAG AAA       384
Asp His Ser Cys Gly His Arg Phe Leu Gly Arg Ser Leu Lys Gln Lys
        115                 120                 125

CTG CAA GAT GCG GTG GGG CAG TGT TTT CCA ATA AAG AAT TGT AGT GGC       432
Leu Gln Asp Ala Val Gly Gln Cys Phe Pro Ile Lys Asn Cys Ser Gly
    130                 135                 140

CGA CAC TCT CCA GGG CTT CCA TCT AAA AGA AAG ATT CAT ATC AGT GAA       480
Arg His Ser Pro Gly Leu Pro Ser Lys Arg Lys Ile His Ile Ser Glu
145                 150                 155                 160

CTC ATG TTA GAT ACG TGC CCC TTC CCA CCT CGC TCA GAT TTA GCC TTT       528
Leu Met Leu Asp Thr Cys Pro Phe Pro Pro Arg Ser Asp Leu Ala Phe
                165                 170                 175

AGG TGG CAT TTT ATT AAA CGA CAC ACT GTT CCT ATG AGT CCC AAC TCA       576
Arg Trp His Phe Ile Lys Arg His Thr Val Pro Met Ser Pro Asn Ser
            180                 185                 190

GAT GAA TGG GTG AGT GCA GAC CTG TCT GAG AGG AAA CTG AGA GAT GCT       624
Asp Glu Trp Val Ser Ala Asp Leu Ser Glu Arg Lys Leu Arg Asp Ala
        195                 200                 205

CAG CTG AAA CGA AGA AAC ACA GAA GAT GAC ATA CCC TGT TTC TCA CAT       672
Gln Leu Lys Arg Arg Asn Thr Glu Asp Asp Ile Pro Cys Phe Ser His
    210                 215                 220

ACC AAT GGC CAG CCT TGT GTC ATA ACT GCC AAC AGT GCT TCG TGT ACA       720
Thr Asn Gly Gln Pro Cys Val Ile Thr Ala Asn Ser Ala Ser Cys Thr
```

```
225                 230                 235                 240

GGT GGT CAC ATA ACT GGT TCT ATG ATG AAC TTG GTC ACA AAC AAC AGC      768
Gly Gly His Ile Thr Gly Ser Met Met Asn Leu Val Thr Asn Asn Ser
                245                 250                 255

ATA GAA GAC AGT GAC ATG GAT TCA GAG GAT GAA ATT ATA ACG CTG TGC      816
Ile Glu Asp Ser Asp Met Asp Ser Glu Asp Glu Ile Ile Thr Leu Cys
            260                 265                 270

ACA AGC TCC AGA AAA AGG AAT AAG CCC AGG TGG GAA ATG GAA GAG GAG      864
Thr Ser Ser Arg Lys Arg Asn Lys Pro Arg Trp Glu Met Glu Glu Glu
        275                 280                 285

ATC CTG CAG TTG GAG GCA CCT CCT AAG TTC CAC ACC CAG ATC GAC TAC      912
Ile Leu Gln Leu Glu Ala Pro Pro Lys Phe His Thr Gln Ile Asp Tyr
    290                 295                 300

GTC CAC TGC CTT GTT CCA GAC CTC CTT CAG ATC AGT AAC AAT CCG TGC      960
Val His Cys Leu Val Pro Asp Leu Leu Gln Ile Ser Asn Asn Pro Cys
305                 310                 315                 320

TAC TGG GGT GTC ATG GAC AAA TAT GCA GCC GAA GCT CTG CTG GAA GGA     1008
Tyr Trp Gly Val Met Asp Lys Tyr Ala Ala Glu Ala Leu Leu Glu Gly
                325                 330                 335

AAG CCA GAG GGC ACC TTT TTA CTT CGA GAT TCA GCG CAG GAA GAT TAT     1056
Lys Pro Glu Gly Thr Phe Leu Leu Arg Asp Ser Ala Gln Glu Asp Tyr
            340                 345                 350

TTA TTC TCT GTT AGT TTT AGA CGC TAC AGT CGT TCT CTT CAT GCT AGA     1104
Leu Phe Ser Val Ser Phe Arg Arg Tyr Ser Arg Ser Leu His Ala Arg
        355                 360                 365

ATT GAG CAG TGG AAT CAT AAC TTT AGC TTT GAT GCC CAT GAT CCT TGT     1152
Ile Glu Gln Trp Asn His Asn Phe Ser Phe Asp Ala His Asp Pro Cys
    370                 375                 380

GTC TTC CAT TCT CCT GAT ATT ACT GGG CTC CTG GAA CAC TAT AAG GAC     1200
Val Phe His Ser Pro Asp Ile Thr Gly Leu Leu Glu His Tyr Lys Asp
385                 390                 395                 400

CCC AGT GCC TGT ATG TTC TTT GAG CCG CTC TTG TCC ACT CCC TTA ATC     1248
Pro Ser Ala Cys Met Phe Phe Glu Pro Leu Leu Ser Thr Pro Leu Ile
                405                 410                 415

CGG ACG TTC CCC TTT TCC TTG CAG CAT ATT TGC AGA ACG GTT ATT TGT     1296
Arg Thr Phe Pro Phe Ser Leu Gln His Ile Cys Arg Thr Val Ile Cys
            420                 425                 430

AAT TGT ACG ACT TAC GAT GGC ATC GAT GCC CTT CCC ATT CCT TCG CCT     1344
Asn Cys Thr Thr Tyr Asp Gly Ile Asp Ala Leu Pro Ile Pro Ser Pro
        435                 440                 445

ATG AAA TTG TAT CTG AAG GAA TAC CAT TAT AAA TCA AAA GTT AGG TTA     1392
Met Lys Leu Tyr Leu Lys Glu Tyr His Tyr Lys Ser Lys Val Arg Leu
    450                 455                 460

CTC AGG ATT GAT GTG CCA GAG CAG CAG TGATGCGGAG AGGTTAGAAT           1439
Leu Arg Ile Asp Val Pro Glu Gln Gln
465                 470

GTCCACCGGA GCTTTTGTTC CCTTTAGTGA GGGTTAATTT CGAGCTTGGC GTAATCATGG   1499
```

TCATAGCTGT TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA CATACGAGCC 1559

GGAAGCATAA AGTGTAAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC ATTAATTGGG 1619

TCGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT GCCAGCTGCA TTACTGAATC 1679

CGCCAACTCG CCGGGACAGC GGTTAGCCTA TTGGGCGCTC TTCACTTCCT CGCTCACTGA 1739

CTCCCTCCCT CGGTCCTTCG CTGCTGCTAC CGTCTCCCCC ATCCAAGCGT TATACGCTAT 1799

CCCCAGAACT GGGAAACCCC GAACACCCTC ACAAAGCTCA CTGCTACCGT ACACGCCCTG 1859

CCGGCTTTTC CTCGTCCCCC CACACCCTAA ACAGCCCTCG AGTGCAACCC CGATATACAT 1919

CTCTTCCCTC AACCCCTGCC TCTGTCCCCG CCTCCGACTT CGCTTCCCCG GATTGCTTTC 1979

CCCCCGTAGT CCGTCCTAGT GCGCCGCGCC TTCCACCCTT CCACCCCTAC GTACCCCCAC 2039

CCCCCAAACC CCCCCCCCCT CCGATAAAAA GTCAGCGCCT TCACCCCCCC GATAAAAATG 2099

GTCCCCTACT TTCCAATGTC TCCCCCCCGG CTCTTCTCGC CACCCAACTC ACCTTTCCGG 2159

CACTGCATCC GGTGCTACCC TCCTGTTTCT CCTCCCCCC 2198

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 473 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Gly Gly Gly Asp Gly Gly Arg Arg Ser Asp Ser Ser Ala * Ala Glu
1 5 10 15

Leu Gly Glu Ile Arg Pro Glu Ser Ala Gln Lys Lys Leu Pro Leu Arg
20 25 30

Lys Ala Glu Asn Thr Ile Phe Ile Thr Leu Glu Ile Val Lys Asn Leu
35 40 45

Phe Lys Met Ala Glu Asn Asn Ser Lys Asn Val Asp Val Arg Pro Lys
50 55 60

Thr Ser Arg Ser Arg Ser Ala Asp Arg Lys Asp Gly Tyr Val Trp Ser
65 70 75 80

Gly Lys Lys Leu Ser Trp Ser Lys Lys Ser Glu Ser Cys Ser Glu Ser
85 90 95

Glu Ala Lys Lys Gly Gln Leu Ser Cys Ser Ser Ile Glu Leu Asp Leu
100 105 110

Asp His Ser Cys Gly His Arg Phe Leu Gly Arg Ser Leu Lys Gln Lys
115 120 125

Leu Gln Asp Ala Val Gly Gln Cys Phe Pro Ile Lys Asn Cys Ser Gly
130 135 140

Arg His Ser Pro Gly Leu Pro Ser Lys Arg Lys Ile His Ile Ser Glu

145 150 155 160

Leu Met Leu Asp Thr Cys Pro Phe Pro Pro Arg Ser Asp Leu Ala Phe
165 170 175

Arg Trp His Phe Ile Lys Arg His Thr Val Pro Met Ser Pro Asn Ser
180 185 190

Asp Glu Trp Val Ser Ala Asp Leu Ser Glu Arg Lys Leu Arg Asp Ala
195 200 205

Gln Leu Lys Arg Arg Asn Thr Glu Asp Asp Ile Pro Cys Phe Ser His
210 215 220

Thr Asn Gly Gln Pro Cys Val Ile Thr Ala Asn Ser Ala Ser Cys Thr
225 230 235 240

Gly Gly His Ile Thr Gly Ser Met Met Asn Leu Val Thr Asn Asn Ser
245 250 255

Ile Glu Asp Ser Asp Met Asp Ser Glu Asp Glu Ile Ile Thr Leu Cys
260 265 270

Thr Ser Ser Arg Lys Arg Asn Lys Pro Arg Trp Glu Met Glu Glu Glu
275 280 285

Ile Leu Gln Leu Glu Ala Pro Pro Lys Phe His Thr Gln Ile Asp Tyr
290 295 300

Val His Cys Leu Val Pro Asp Leu Leu Gln Ile Ser Asn Asn Pro Cys
305 310 315 320

Tyr Trp Gly Val Met Asp Lys Tyr Ala Ala Glu Ala Leu Leu Glu Gly
325 330 335

Lys Pro Glu Gly Thr Phe Leu Leu Arg Asp Ser Ala Gln Glu Asp Tyr
340 345 350

Leu Phe Ser Val Ser Phe Arg Arg Tyr Ser Arg Ser Leu His Ala Arg
355 360 365

Ile Glu Gln Trp Asn His Asn Phe Ser Phe Asp Ala His Asp Pro Cys
370 375 380

Val Phe His Ser Pro Asp Ile Thr Gly Leu Leu Glu His Tyr Lys Asp
385 390 395 400

Pro Ser Ala Cys Met Phe Phe Glu Pro Leu Leu Ser Thr Pro Leu Ile
405 410 415

Arg Thr Phe Pro Phe Ser Leu Gln His Ile Cys Arg Thr Val Ile Cys
420 425 430

Asn Cys Thr Thr Tyr Asp Gly Ile Asp Ala Leu Pro Ile Pro Ser Pro
435 440 445

Met Lys Leu Tyr Leu Lys Glu Tyr His Tyr Lys Ser Lys Val Arg Leu
450 455 460

Leu Arg Ile Asp Val Pro Glu Gln Gln
465 470

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2254 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 117..1724

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 740
(D) OTHER INFORMATION: /note= "Nucleotide may be A or C at positions: 740, 797, 2139, and 2184."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 761
(D) OTHER INFORMATION: /note= "Nucleotide may be G or T at positions: 761, 1313, 1508, and 2226."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 746
(D) OTHER INFORMATION: /note= "Nucleotide may be C or T at positions 746, 1460, 1499, 2009, 2010, 2199, and 2225. "

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 788
(D) OTHER INFORMATION: /note= "Nucleotide may be A or G at positions 788, 863, 1550, 2178, 2188, 2197, and 2211."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1163
(D) OTHER INFORMATION: /note= "Nucleotide may be G or C at positions: 1163, and 1544."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 2058
(D) OTHER INFORMATION: /note= "Nucleotide may be A or T at positions 2058, and 2128."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 2251
(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, T, or G at position 2251."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
CTCGGGCCGG GATGGATCCG CCGGGAAGAG GAAGACAAGC GGAGCGTTGA GCCCCTGCGC      60

ACGGTGCCCC GCGCGTAGTG GGAGCTTACT CGCAGTAGCT CTCGCTCTTC TAATCA         116

ATG GAT AAA GTG GGG AAA ATG TGG AAC AAC TTA AAA TAC AGA TGC CAG      164
Met Asp Lys Val Gly Lys Met Trp Asn Asn Leu Lys Tyr Arg Cys Gln
```

```
 1               5                   10                  15

AAT CTC TTC AGC CAC GAG GGA GGA AGC CGT AAT GAG AAC GTG GAG ATG     212
Asn Leu Phe Ser His Glu Gly Gly Ser Arg Asn Glu Asn Val Glu Met
            20                  25                  30

AAC CCC AAC AGA TGT CCG TCT GTC AAA GAG AAA AGC ATC AGT CTG GGA     260
Asn Pro Asn Arg Cys Pro Ser Val Lys Glu Lys Ser Ile Ser Leu Gly
        35                  40                  45

GAG GCA GCT CCC CAG CAA GAG AGC AGT CCC TTA AGA GAA AAT GTT GCC     308
Glu Ala Ala Pro Gln Gln Glu Ser Ser Pro Leu Arg Glu Asn Val Ala
    50                  55                  60

TTA CAG CTG GGA CTG AGC CCT TCC AAG ACC TTT TCC AGG CGG AAC CAA     356
Leu Gln Leu Gly Leu Ser Pro Ser Lys Thr Phe Ser Arg Arg Asn Gln
65                  70                  75                  80

AAC TGT GCC GCA GAG ATC CCT CAA GTG GTT GAA ATC AGC ATC GAG AAA     404
Asn Cys Ala Ala Glu Ile Pro Gln Val Val Glu Ile Ser Ile Glu Lys
                85                  90                  95

GAC AGT GAC TCG GGT GCC ACC CCA GGA ACG AGG CTT GCA CGG AGA GAC     452
Asp Ser Asp Ser Gly Ala Thr Pro Gly Thr Arg Leu Ala Arg Arg Asp
            100                 105                 110

TCC TAC TCG CGG CAC GCC CCG TGG GGA GGA AAG AAG AAA CAT TCC TGT     500
Ser Tyr Ser Arg His Ala Pro Trp Gly Gly Lys Lys Lys His Ser Cys
        115                 120                 125

TCC ACA AAG ACC CAG AGT TCA TTG GAT ACC GAG AAA AAG TTT GGT AGA     548
Ser Thr Lys Thr Gln Ser Ser Leu Asp Thr Glu Lys Lys Phe Gly Arg
    130                 135                 140

ACT CGA AGC GGC CTT CAG AGG CGA GAG CGG CGC TAT GGA GTC AGC TCC     596
Thr Arg Ser Gly Leu Gln Arg Arg Glu Arg Arg Tyr Gly Val Ser Ser
145                 150                 155                 160

ATG CAG GAC ATG GAC AGC GTT TCT AGC CGC GCG GTC GGG AGC CGC TCC     644
Met Gln Asp Met Asp Ser Val Ser Ser Arg Ala Val Gly Ser Arg Ser
                165                 170                 175

CTG AGG CAG AGG CTC CAG GAC ACG GTG GGT TTG TGT TTT CCC ATG AGA     692
Leu Arg Gln Arg Leu Gln Asp Thr Val Gly Leu Cys Phe Pro Met Arg
            180                 185                 190

ACT TAC AGC AAG CAG TCA AAG CCA CTC TTT TCC AAT AAA AGA AAA ATC     740
Thr Tyr Ser Lys Gln Ser Lys Pro Leu Phe Ser Asn Lys Arg Lys Ile
        195                 200                 205

CAT CTC TCT GAA TTA ATG CTG GAG AAA TGC CCT TTT CCT GCT GGC TCG     788
His Leu Ser Glu Leu Met Leu Glu Lys Cys Pro Phe Pro Ala Gly Ser
    210                 215                 220

GAT TTA GCC CAA AAG TGG CAT TTG ATT AAA CAG CAT ACA GCT CCT GTG     836
Asp Leu Ala Gln Lys Trp His Leu Ile Lys Gln His Thr Ala Pro Val
225                 230                 235                 240

AGC CCA CAT TCA ACA TTT TTT GAT ACG TTT GAT CCA TCT TTG GTT TCT     884
Ser Pro His Ser Thr Phe Phe Asp Thr Phe Asp Pro Ser Leu Val Ser
                245                 250                 255

ACA GAA GAT GAA GAA GAT AGG CTT AGA GAG AGA AGG CGG CTT AGT ATT     932
Thr Glu Asp Glu Glu Asp Arg Leu Arg Glu Arg Arg Arg Leu Ser Ile
```

```
        260                 265                 270

GAA GAA GGG GTT GAT CCC CCT CCC AAT GCA CAA ATA CAT ACA TTT GAA      980
Glu Glu Gly Val Asp Pro Pro Pro Asn Ala Gln Ile His Thr Phe Glu
        275                 280                 285

GCT ACT GCA CAG GTT AAT CCA TTA TTT AAA CTG GGA CCA AAA TTA GCT     1028
Ala Thr Ala Gln Val Asn Pro Leu Phe Lys Leu Gly Pro Lys Leu Ala
    290                 295                 300

CCT GGA ATG ACT GAA ATA AGT GGG GAC AGT TCT GCA ATT CCA CAA GCT     1076
Pro Gly Met Thr Glu Ile Ser Gly Asp Ser Ser Ala Ile Pro Gln Ala
305                 310                 315                 320

AAT TGT GAC TCG GAA GAG GAT ACA ACC ACC CTG TGT TTG CAG TCA CGG     1124
Asn Cys Asp Ser Glu Glu Asp Thr Thr Thr Leu Cys Leu Gln Ser Arg
                325                 330                 335

AGG CAG AAG CAG CGT CAG ATA TCT GGA GAC AGC CAT ACC CAT GTT AGC     1172
Arg Gln Lys Gln Arg Gln Ile Ser Gly Asp Ser His Thr His Val Ser
            340                 345                 350

AGA CAG GGA GCT TGG AAA GTC CAC ACA CAG ATT GAT TAC ATA CAC TGC     1220
Arg Gln Gly Ala Trp Lys Val His Thr Gln Ile Asp Tyr Ile His Cys
        355                 360                 365

CTC GTG CCT GAT TTG CTT CAA ATT ACA GGG AAT CCC TGT TAC TGG GGA     1268
Leu Val Pro Asp Leu Leu Gln Ile Thr Gly Asn Pro Cys Tyr Trp Gly
    370                 375                 380

GTG ATG GAC CGT TAT GAA GCA GAA GCC CTC TCC GAA GGG AAA CCG GAA     1316
Val Met Asp Arg Tyr Glu Ala Glu Ala Leu Ser Glu Gly Lys Pro Glu
385                 390                 395                 400

GGC ACG TTC TTG CTC AGG GAC TCT GCA CAG GAG GAC TAC CTC TTC TCT     1364
Gly Thr Phe Leu Leu Arg Asp Ser Ala Gln Glu Asp Tyr Leu Phe Ser
                405                 410                 415

GTG AGT TCC GCC GCT ACA ACA GGA TCT CTG CAC GCC CGG ATC GAG CAG     1412
Val Ser Ser Ala Ala Thr Thr Gly Ser Leu His Ala Arg Ile Glu Gln
            420                 425                 430

TGG AAC CAC AAC TTC AGC TTC GAT GCC CAT GAC CCC TGC GTG TTT CAC     1460
Trp Asn His Asn Phe Ser Phe Asp Ala His Asp Pro Cys Val Phe His
        435                 440                 445

TCC TCC ACT GTC ACG GGG CTT CTC GAA CAC TAT AAA GAC CCC AGT TCG     1508
Ser Ser Thr Val Thr Gly Leu Leu Glu His Tyr Lys Asp Pro Ser Ser
    450                 455                 460

TGC ATG TTT TTT GAA CCG TTG CTA ACG ATA TCA CTC AAT AGG ACT TTC     1556
Cys Met Phe Phe Glu Pro Leu Leu Thr Ile Ser Leu Asn Arg Thr Phe
465                 470                 475                 480

CCT TTC AGC CTG CAG TAT ATC TGC CGC GCA GTG ATC TGC AGA TGC ACT     1604
Pro Phe Ser Leu Gln Tyr Ile Cys Arg Ala Val Ile Cys Arg Cys Thr
                485                 490                 495

ACG TAT GAT GGG ATT GAC GGG CTC CCG CTA CCG TCG ATG TTA CAG GAT     1652
Thr Tyr Asp Gly Ile Asp Gly Leu Pro Leu Pro Ser Met Leu Gln Asp
            500                 505                 510

TTT TTA AAA GAG TAT CAT TAT AAA CAA AAA GTT AGA GTT CGC TGG TTG     1700
Phe Leu Lys Glu Tyr His Tyr Lys Gln Lys Val Arg Val Arg Trp Leu
```

515 520 525

GAA CGA GAA CCA GTC AAG GCA AAG TAAACTCTCC GGTCCCCAAA GGGTGTTAAC 1754
Glu Arg Glu Pro Val Lys Ala Lys
530 535

TAGGTCCGCT TTCATGTGCA TCAGACAGTA CACCTATAGC AAGCACACGT AGCAGTGTTA 1814

GGCTTTTTCA TACAGTATGT AAGCTTAGTG TTAGTATCTG TCAGATGCTA CCTGCTGTTA 1874

CTTATTCAGA TAAACATGGT GCCTATTGGA ACAATAGCGG ATAGAGCTAC AGGTGTTCAG 1934

TAAGACTACA AAAACATTTT GCCTATTTCG CTAACAGTTT GGTTTTTAAT GGCTGTGGTA 1994

TTTGAGTGAG GCAACCCTGG GGCATTTGTT ATGAAGAATT CTATTTCTTA CTGAAGAACA 2054

AATAATTAAT ATTGGATGAG TATTTCAACA GTGTGACTAA TGTTTGAAAT TATTTTTTCC 2114

TAAGAGTTTT TCCTATAACC TTCCAAAAGT CGTGATGTTT GTAGTTACCA TAATCCAGCT 2174

TTGAAGTCCA AAAGGATTAA AGGCCGCCTC CCTTTGAAAA ATGCCATTTC CGGCCCCAAG 2234

GCCTAGTGCC GTCCCTCCGG 2254

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 536 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Asp Lys Val Gly Lys Met Trp Asn Asn Leu Lys Tyr Arg Cys Gln
1 5 10 15

Asn Leu Phe Ser His Glu Gly Gly Ser Arg Asn Glu Asn Val Glu Met
20 25 30

Asn Pro Asn Arg Cys Pro Ser Val Lys Glu Lys Ser Ile Ser Leu Gly
35 40 45

Glu Ala Ala Pro Gln Gln Glu Ser Ser Pro Leu Arg Glu Asn Val Ala
50 55 60

Leu Gln Leu Gly Leu Ser Pro Ser Lys Thr Phe Ser Arg Arg Asn Gln
65 70 75 80

Asn Cys Ala Ala Glu Ile Pro Gln Val Val Glu Ile Ser Ile Glu Lys
85 90 95

Asp Ser Asp Ser Gly Ala Thr Pro Gly Thr Arg Leu Ala Arg Arg Asp
100 105 110

Ser Tyr Ser Arg His Ala Pro Trp Gly Gly Lys Lys Lys His Ser Cys
115 120 125

Ser Thr Lys Thr Gln Ser Ser Leu Asp Thr Glu Lys Lys Phe Gly Arg
130 135 140

Thr Arg Ser Gly Leu Gln Arg Arg Glu Arg Arg Tyr Gly Val Ser Ser

```
145                 150                 155                 160

Met Gln Asp Met Asp Ser Val Ser Ser Arg Ala Val Gly Ser Arg Ser
                165                 170                 175

Leu Arg Gln Arg Leu Gln Asp Thr Val Gly Leu Cys Phe Pro Met Arg
            180                 185                 190

Thr Tyr Ser Lys Gln Ser Lys Pro Leu Phe Ser Asn Lys Arg Lys Ile
        195                 200                 205

His Leu Ser Glu Leu Met Leu Glu Lys Cys Pro Phe Pro Ala Gly Ser
    210                 215                 220

Asp Leu Ala Gln Lys Trp His Leu Ile Lys Gln His Thr Ala Pro Val
225                 230                 235                 240

Ser Pro His Ser Thr Phe Phe Asp Thr Phe Asp Pro Ser Leu Val Ser
                245                 250                 255

Thr Glu Asp Glu Glu Asp Arg Leu Arg Glu Arg Arg Arg Leu Ser Ile
            260                 265                 270

Glu Glu Gly Val Asp Pro Pro Pro Asn Ala Gln Ile His Thr Phe Glu
        275                 280                 285

Ala Thr Ala Gln Val Asn Pro Leu Phe Lys Leu Gly Pro Lys Leu Ala
    290                 295                 300

Pro Gly Met Thr Glu Ile Ser Gly Asp Ser Ser Ala Ile Pro Gln Ala
305                 310                 315                 320

Asn Cys Asp Ser Glu Glu Asp Thr Thr Thr Leu Cys Leu Gln Ser Arg
                325                 330                 335

Arg Gln Lys Gln Arg Gln Ile Ser Gly Asp Ser His Thr His Val Ser
            340                 345                 350

Arg Gln Gly Ala Trp Lys Val His Thr Gln Ile Asp Tyr Ile His Cys
        355                 360                 365

Leu Val Pro Asp Leu Leu Gln Ile Thr Gly Asn Pro Cys Tyr Trp Gly
    370                 375                 380

Val Met Asp Arg Tyr Glu Ala Glu Ala Leu Ser Glu Gly Lys Pro Glu
385                 390                 395                 400

Gly Thr Phe Leu Leu Arg Asp Ser Ala Gln Glu Asp Tyr Leu Phe Ser
                405                 410                 415

Val Ser Ser Ala Ala Thr Thr Gly Ser Leu His Ala Arg Ile Glu Gln
            420                 425                 430

Trp Asn His Asn Phe Ser Phe Asp Ala His Asp Pro Cys Val Phe His
        435                 440                 445

Ser Ser Thr Val Thr Gly Leu Leu Glu His Tyr Lys Asp Pro Ser Ser
    450                 455                 460

Cys Met Phe Phe Glu Pro Leu Leu Thr Ile Ser Leu Asn Arg Thr Phe
465                 470                 475                 480

Pro Phe Ser Leu Gln Tyr Ile Cys Arg Ala Val Ile Cys Arg Cys Thr
                485                 490                 495
```

```
Thr Tyr Asp Gly Ile Asp Gly Leu Pro Leu Pro Ser Met Leu Gln Asp
            500                 505                 510

Phe Leu Lys Glu Tyr His Tyr Lys Gln Lys Val Arg Val Arg Trp Leu
        515                 520                 525

Glu Arg Glu Pro Val Lys Ala Lys
    530                 535
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2206 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..1375

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 2078
(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, T, or G at positions: 2078, and 2116."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 2063
(D) OTHER INFORMATION: /note= "Nucleotide may be G or C at position 2063."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
G GAG CGC GGC CTG GAG ACT AAC AGC TGC TCG GAA GAG GAG CTC AGC      46
  Glu Arg Gly Leu Glu Thr Asn Ser Cys Ser Glu Glu Glu Leu Ser
    1               5                  10                  15

AGC CCG GGT CGC GGA GGA GGA GGG GGC GGC CGG CTT CTG CTG CAG CCC    94
Ser Pro Gly Arg Gly Gly Gly Gly Gly Gly Arg Leu Leu Leu Gln Pro
                 20                  25                  30

CCA GGC CCT GAA TTA CCT CCG GTG CCC TTC CCG CTG CAG GAC TTG GTC   142
Pro Gly Pro Glu Leu Pro Pro Val Pro Phe Pro Leu Gln Asp Leu Val
             35                  40                  45

CCT CTG GGG CGC CTG AGT AGA GGG GAG CAG CAG CAG CAG CAG CAG CAG   190
Pro Leu Gly Arg Leu Ser Arg Gly Glu Gln Gln Gln Gln Gln Gln Gln
         50                  55                  60

CAA CCT CCC CCG CCC CCG CCT CCT CCC GGG CCC CTC CGG CCA CTC GCG   238
Gln Pro Pro Pro Pro Pro Pro Pro Pro Gly Pro Leu Arg Pro Leu Ala
     65                  70                  75

GGT CCT TCT CGG AAG GGC TCC TTC AAA ATC CGC CTC AGT CGC CTC TTT   286
Gly Pro Ser Arg Lys Gly Ser Phe Lys Ile Arg Leu Ser Arg Leu Phe
 80                  85                  90                  95
```

```
CGC ACC AAG AGC TGC AAC GGT GGC TCC GGC GGT GGG GAT GGG ACC GGC      334
Arg Thr Lys Ser Cys Asn Gly Gly Ser Gly Gly Gly Asp Gly Thr Gly
                100                 105                 110

AAG AGG CCT TCT GGA GAG CTG GCT GCT TCA GCT GCG AGC CTG ACA GAC      382
Lys Arg Pro Ser Gly Glu Leu Ala Ala Ser Ala Ala Ser Leu Thr Asp
            115                 120                 125

ATG GGA GGC TCT GCG GGC CGG GAG CTG GAC GCG GGG AGG AAA CCC AAG      430
Met Gly Gly Ser Ala Gly Arg Glu Leu Asp Ala Gly Arg Lys Pro Lys
        130                 135                 140

TTG ACA AGA ACT CAA AGT GCC TTT TCT CCG GTC TCC TTC AGC CCC CTG      478
Leu Thr Arg Thr Gln Ser Ala Phe Ser Pro Val Ser Phe Ser Pro Leu
    145                 150                 155

TTC ACA GGT GAA ACT GTG TCG CTT GTG GAT GTG GAC ATT TCT CAG CGG      526
Phe Thr Gly Glu Thr Val Ser Leu Val Asp Val Asp Ile Ser Gln Arg
160                 165                 170                 175

GGC CTG ACC TCT CCA CAC CCT CCA ACT CCC CCT CCT CCT CCG AGA AGA      574
Gly Leu Thr Ser Pro His Pro Pro Thr Pro Pro Pro Pro Pro Arg Arg
                180                 185                 190

AGC CTC AGC CTC CTA GAT GAT ATC AGT GGG ACG CTG CCT ACA TCT GTC      622
Ser Leu Ser Leu Leu Asp Asp Ile Ser Gly Thr Leu Pro Thr Ser Val
            195                 200                 205

CTT GTG GCT CCG ATG GGG TCT TCC TTG CAG TCT TTC CCC CTA CCT CCG      670
Leu Val Ala Pro Met Gly Ser Ser Leu Gln Ser Phe Pro Leu Pro Pro
        210                 215                 220

CCT CCT CCA CCC CAT GCC CCA GAT GCA TTT CCC CGG ATT GCT CCC ATC      718
Pro Pro Pro Pro His Ala Pro Asp Ala Phe Pro Arg Ile Ala Pro Ile
    225                 230                 235

CGA GCA GCT GAA TCC CTG CAC AGC CAA CCC CCA CAG CAC CTC CAG TGT      766
Arg Ala Ala Glu Ser Leu His Ser Gln Pro Pro Gln His Leu Gln Cys
240                 245                 250                 255

CCC CTC TAC CGG CCT GAC TCG AGC AGC TTT GCA GCC AGC CTT CGA GAG      814
Pro Leu Tyr Arg Pro Asp Ser Ser Ser Phe Ala Ala Ser Leu Arg Glu
                260                 265                 270

TTG GAG AAG TGT GGT TGG TAT TGG GGG CCA ATG AAT TGG GAA GAT GCA      862
Leu Glu Lys Cys Gly Trp Tyr Trp Gly Pro Met Asn Trp Glu Asp Ala
            275                 280                 285

GAG ATG AAG CTG AAA GGG AAA CCA GAT GGT TCT TTC CTG GTA CGA GAC      910
Glu Met Lys Leu Lys Gly Lys Pro Asp Gly Ser Phe Leu Val Arg Asp
        290                 295                 300

AGT TCT GAT CCT CGT TAC ATC CTG AGC CTC AGT TTC CGA TCA CAG GGT      958
Ser Ser Asp Pro Arg Tyr Ile Leu Ser Leu Ser Phe Arg Ser Gln Gly
    305                 310                 315

ATC ACC CAC CAC ACT AGA ATG GAG CAC TAC AGA GGA ACC TTC AGC CTG     1006
Ile Thr His His Thr Arg Met Glu His Tyr Arg Gly Thr Phe Ser Leu
320                 325                 330                 335

TGG TGT CAT CCC AAG TTT GAG GAC CGC TGT CAA TCT GTT GTA GAG TTT     1054
Trp Cys His Pro Lys Phe Glu Asp Arg Cys Gln Ser Val Val Glu Phe
                340                 345                 350
```

```
ATT AAG AGA GCC ATT ATG CAC TCC AAG AAT GGA AAG TTT CTC TAT TTC      1102
Ile Lys Arg Ala Ile Met His Ser Lys Asn Gly Lys Phe Leu Tyr Phe
            355                 360                 365

TTA AGA TCC AGG GTT CCA GGA CTG CCA CCA ACT CCT GTC CAG CTG CTC      1150
Leu Arg Ser Arg Val Pro Gly Leu Pro Pro Thr Pro Val Gln Leu Leu
        370                 375                 380

TAT CCA GTG TCC CGA TTC AGC AAT GTC AAA TCC CTC CAG CAC CTT TGC      1198
Tyr Pro Val Ser Arg Phe Ser Asn Val Lys Ser Leu Gln His Leu Cys
    385                 390                 395

AGA TTC CGG ATA CGA CAG CTC GTC AGG ATA GAT CAC ATC CCA GAT CTC      1246
Arg Phe Arg Ile Arg Gln Leu Val Arg Ile Asp His Ile Pro Asp Leu
400                 405                 410                 415

CCA CTG CCT AAA CCT CTG ATC TCT TAT ATC CGA AAG TTC TAC TAC TAT      1294
Pro Leu Pro Lys Pro Leu Ile Ser Tyr Ile Arg Lys Phe Tyr Tyr Tyr
                420                 425                 430

GAT CCT CAG GAA GAG GTA TAC CTG TCT CTA AAG GAA GCG CAG CTC ATT      1342
Asp Pro Gln Glu Glu Val Tyr Leu Ser Leu Lys Glu Ala Gln Leu Ile
            435                 440                 445

TCC AAA CAG AAG CAA GAG GTG GAA CCC TCC ACG TAGCGAGGGG CTCCCTGCTG    1395
Ser Lys Gln Lys Gln Glu Val Glu Pro Ser Thr
        450                 455

GTCACCACCA AGGGCATTTG GTTGCCAAGC TCCAGCTTTG AAGAACCAAA TTAAGCTACC    1455

ATGAAAAGAA GAGGAAAAGT GAGGGAACAG GAAGGTTGGG ATTCTCTGTG CAGAGACTTT    1515

GGTTCCCCAC GCAGCCCTGG GGCTTGGAAG AAGCACATGA CCGTACTCTG CGTGGGGCTC    1575

CACCTCACAC CCACCCCTGG GCATCTTAGG ACTGGAGGGG CTCCTTGGAA AACTGGAAGA    1635

AGTCTCAACA CTGTTTCTTT TTCAAAAAAA AAAAAAAAAA AGATGCGGCC GCAAGCTTAT    1695

TCCCTTTAGT GAGGGTTAAT TTTAGCTTGG CACTGGCCGT CGTTTTACAA CGTCGTGACT    1755

GGGAAAACCC TGGCGTTACC CAACTTAATC GCCTTGCAGC ACATCCCCCT TTCGCCAGCT    1815

GGCGTAATAG CGAAGAGGCC CGCACCGATC GCCCTTCCCA ACAGTTGCGC AGCCTGAATG    1875

GCGAATGGGA CGCGCCCTGT AGCGGCGCAT TAACGCGCGG CGGGTGTGGT GGTTACGCGC    1935

AGCGTGACCG CTACACTTGC CAGCGCCCTA CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT    1995

TTCTCGCCAC GTTCGCCGGC TTTCCCCGTC AACTCTAAAT CGGGGGCTCC CTTTAGGTTC    2055

CGATTTACTG CTTTACGCAC TCCACCCCAA AACTTGATTA GGTGATGTCA CTTATGGCAC    2115

CCCTGATAAC GTTTCCCCTT ACTTTGATCA CTTCTTTATA TGATCTTTCC AATGAAACAT    2175

CACCTACTCG TCATCTTTAT TTAAAGATTT G                                   2206
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 458 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Glu Arg Gly Leu Glu Thr Asn Ser Cys Ser Glu Glu Glu Leu Ser Ser
 1               5                  10                  15

Pro Gly Arg Gly Gly Gly Gly Gly Gly Arg Leu Leu Leu Gln Pro Pro
            20                  25                  30

Gly Pro Glu Leu Pro Pro Val Pro Phe Pro Leu Gln Asp Leu Val Pro
        35                  40                  45

Leu Gly Arg Leu Ser Arg Gly Glu Gln Gln Gln Gln Gln Gln Gln Gln
    50                  55                  60

Pro Pro Pro Pro Pro Pro Pro Pro Gly Pro Leu Arg Pro Leu Ala Gly
65                  70                  75                  80

Pro Ser Arg Lys Gly Ser Phe Lys Ile Arg Leu Ser Arg Leu Phe Arg
                85                  90                  95

Thr Lys Ser Cys Asn Gly Gly Ser Gly Gly Gly Asp Gly Thr Gly Lys
            100                 105                 110

Arg Pro Ser Gly Glu Leu Ala Ala Ser Ala Ala Ser Leu Thr Asp Met
        115                 120                 125

Gly Gly Ser Ala Gly Arg Glu Leu Asp Ala Gly Arg Lys Pro Lys Leu
    130                 135                 140

Thr Arg Thr Gln Ser Ala Phe Ser Pro Val Ser Phe Ser Pro Leu Phe
145                 150                 155                 160

Thr Gly Glu Thr Val Ser Leu Val Asp Val Asp Ile Ser Gln Arg Gly
                165                 170                 175

Leu Thr Ser Pro His Pro Pro Thr Pro Pro Pro Pro Pro Arg Arg Ser
            180                 185                 190

Leu Ser Leu Leu Asp Asp Ile Ser Gly Thr Leu Pro Thr Ser Val Leu
        195                 200                 205

Val Ala Pro Met Gly Ser Ser Leu Gln Ser Phe Pro Leu Pro Pro Pro
    210                 215                 220

Pro Pro Pro His Ala Pro Asp Ala Phe Pro Arg Ile Ala Pro Ile Arg
225                 230                 235                 240

Ala Ala Glu Ser Leu His Ser Gln Pro Pro Gln His Leu Gln Cys Pro
                245                 250                 255

Leu Tyr Arg Pro Asp Ser Ser Ser Phe Ala Ala Ser Leu Arg Glu Leu
            260                 265                 270

Glu Lys Cys Gly Trp Tyr Trp Gly Pro Met Asn Trp Glu Asp Ala Glu
        275                 280                 285

Met Lys Leu Lys Gly Lys Pro Asp Gly Ser Phe Leu Val Arg Asp Ser
    290                 295                 300

Ser Asp Pro Arg Tyr Ile Leu Ser Leu Ser Phe Arg Ser Gln Gly Ile
305                 310                 315                 320
```

```
Thr His His Thr Arg Met Glu His Tyr Arg Gly Thr Phe Ser Leu Trp
                325                 330                 335

Cys His Pro Lys Phe Glu Asp Arg Cys Gln Ser Val Val Glu Phe Ile
            340                 345                 350

Lys Arg Ala Ile Met His Ser Lys Asn Gly Lys Phe Leu Tyr Phe Leu
        355                 360                 365

Arg Ser Arg Val Pro Gly Leu Pro Pro Thr Pro Val Gln Leu Leu Tyr
    370                 375                 380

Pro Val Ser Arg Phe Ser Asn Val Lys Ser Leu Gln His Leu Cys Arg
385                 390                 395                 400

Phe Arg Ile Arg Gln Leu Val Arg Ile Asp His Ile Pro Asp Leu Pro
                405                 410                 415

Leu Pro Lys Pro Leu Ile Ser Tyr Ile Arg Lys Phe Tyr Tyr Tyr Asp
            420                 425                 430

Pro Gln Glu Glu Val Tyr Leu Ser Leu Lys Glu Ala Gln Leu Ile Ser
        435                 440                 445

Lys Gln Lys Gln Glu Val Glu Pro Ser Thr
    450                 455
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1390 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 453..1388

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 108
(D) OTHER INFORMATION: /note= "Nucleotide may be A, C, T, or G at positions: 108, and 109."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 236
(D) OTHER INFORMATION: /note= "Nucleotide may be A or G at positions: 236, 238, and 1258."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 233
(D) OTHER INFORMATION: /note= "Nucleotide may be G or T at position 233."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 234
(D) OTHER INFORMATION: /note= "Nucleotide may be G or C at

position 234."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 237
(D) OTHER INFORMATION: /note= "Nucleotide may be C or T at position 237."

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 239
(D) OTHER INFORMATION: /note= "Nucleotide may be A or T at position 239."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
CGGACGCGTG GGTTTGGCTG TGAATATTCT ATTTGCTTGC AGTATCTGTT TCTCTTCCTA      60

GGCTCAAGTT GGTGACCCAA GCCTATTGTA AACAAGTGAT TATCTCACCG GGAGATGCCA     120

ATGGAGTAAC AATTTGTTAA CCTTACGTTT TCTGTCTGTA TATTTTTTTA AAAATCTGGT     180

AGTTTCTGGA AAAAAAAGAG AAGGGGGTTT GTAGTACTTA ACCCTATTTA TTGCCACGAG     240

TTTTAGTTAA TTAGTTTTTG GAATAAATGG ATTTCAGTAT AGCTTTGTGG TTAAATTGCA     300

TTGCCTTTAT TTTATGTTTA GGCTTATTTT TAAATTAACA TTTAACAGAA ACATTTGAAA     360

TAGAATTTGC ATGTCTGCCT TAATTAACTT AAAGACTGAT TTTAATCTGA CTATGACACT     420

GAGCATATTC TTTAAATTAC TCATAATTTA TA ATG CTT AAT ATA ATC TTA ATT       473
                                    Met Leu Asn Ile Ile Leu Ile
                                     1               5

AAA TTT AGC AGT TTT AGT ATA AGA TGT GCC ATT TTG TCC TCT GTA TGT       521
Lys Phe Ser Ser Phe Ser Ile Arg Cys Ala Ile Leu Ser Ser Val Cys
        10                  15                  20

CTG AAT GAA GCT ATA ACA TTT GCC TTT TTA TTG CAG GTT TTC CTT TGG       569
Leu Asn Glu Ala Ile Thr Phe Ala Phe Leu Leu Gln Val Phe Leu Trp
    25                  30                  35

AAT ATG GAT AAA TAC ACC ATG ATA CGG AAA CTA GAA GGA CAT CAC CAT       617
Asn Met Asp Lys Tyr Thr Met Ile Arg Lys Leu Glu Gly His His His
40                  45                  50                  55

GAT GTG GTA GCT TGT GAC TTT TCT CCT GAT GGA GCA TTA CTG GCT ACT       665
Asp Val Val Ala Cys Asp Phe Ser Pro Asp Gly Ala Leu Leu Ala Thr
                60                  65                  70

GCA TCT TAT GAT ACT CGA GTA TAT ATC TGG GAT CCA CAT AAT GGA GAC       713
Ala Ser Tyr Asp Thr Arg Val Tyr Ile Trp Asp Pro His Asn Gly Asp
            75                  80                  85

ATT CTG ATG GAA TTT GGG CAC CTG TTT CCC CCA CCT ACT CCA ATA TTT       761
Ile Leu Met Glu Phe Gly His Leu Phe Pro Pro Pro Thr Pro Ile Phe
        90                  95                 100

GCT GGA GGA GCA AAT GAC CGG TGG GTA CGA TCT GTA TCT TTT AGC CAT       809
Ala Gly Gly Ala Asn Asp Arg Trp Val Arg Ser Val Ser Phe Ser His
    105                 110                 115

GAT GGA CTG CAT GTT GCA AGC CTT GCT GAT GAT AAA ATG GTG AGG TTC       857
```

```
Asp Gly Leu His Val Ala Ser Leu Ala Asp Asp Lys Met Val Arg Phe
120                 125                 130                 135

TGG AGA ATT GAT GAG GAT TAT CCA GTG CAA GTT GCA CCT TTG AGC AAT       905
Trp Arg Ile Asp Glu Asp Tyr Pro Val Gln Val Ala Pro Leu Ser Asn
                140                 145                 150

GGT CTT TGC TGT GCC TTC TCT ACT GAT GGC AGT GTT TTA GCT GCT GGG       953
Gly Leu Cys Cys Ala Phe Ser Thr Asp Gly Ser Val Leu Ala Ala Gly
            155                 160                 165

ACA CAT GAC GGA AGT GTG TAT TTT TGG GCC ACT CCA CGG CAG GTC CCT      1001
Thr His Asp Gly Ser Val Tyr Phe Trp Ala Thr Pro Arg Gln Val Pro
        170                 175                 180

AGC CTG CAA CAT TTA TGT CGC ATG TCA ATC CGA AGA GTG ATG CCC ACC      1049
Ser Leu Gln His Leu Cys Arg Met Ser Ile Arg Arg Val Met Pro Thr
    185                 190                 195

CAA GAA GTT CAG GAG CTG CCG ATT CCT TCC AAG CTT TTG GAG TTT CTC      1097
Gln Glu Val Gln Glu Leu Pro Ile Pro Ser Lys Leu Leu Glu Phe Leu
200                 205                 210                 215

TCG TAT CGT ATT TAG AAG ATT CTG CCT TCC CTA GTA GTA GGG ACT GAC      1145
Ser Tyr Arg Ile  *  Lys Ile Leu Pro Ser Leu Val Val Gly Thr Asp
                220                 225                 230

AGA ATA CAC TTA ACA CAA ACC TCA AGC TTT ACT GAC TTC AAT TAT CTG      1193
Arg Ile His Leu Thr Gln Thr Ser Ser Phe Thr Asp Phe Asn Tyr Leu
            235                 240                 245

TTT TTA AAG ACG TAG AAG ATT TAT TTA ATT TGA TAT GTT CTT GTA CTG      1241
Phe Leu Lys Thr  *  Lys Ile Tyr Leu Ile  *  Tyr Val Leu Val Leu
        250                 255                 260

CAT TTT GAT CAG TTG AAG CTT TTA AAA TAT TAT TTA TAG ACA ATA GAA      1289
His Phe Asp Gln Leu Lys Leu Leu Lys Tyr Tyr Leu  *  Thr Ile Glu
    265                 270                 275

GTA TTT CTG AAC ATA TCA AAT ATA AAT TTT TTT AAA GAT CTA ACT GTG      1337
Val Phe Leu Asn Ile Ser Asn Ile Asn Phe Phe Lys Asp Leu Thr Val
280                 285                 290                 295

AAA AAC ATA CAT ACC TGT ACA TAT TTA GAT ATA AGC TGC TAT ATG TTG      1385
Lys Asn Ile His Thr Cys Thr Tyr Leu Asp Ile Ser Cys Tyr Met Leu
                300                 305                 310

AAT GG                                                               1390
Asn
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 312 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Met Leu Asn Ile Ile Leu Ile Lys Phe Ser Ser Phe Ser Ile Arg Cys
```

1 5 10 15

Ala Ile Leu Ser Ser Val Cys Leu Asn Glu Ala Ile Thr Phe Ala Phe
20 25 30

Leu Leu Gln Val Phe Leu Trp Asn Met Asp Lys Tyr Thr Met Ile Arg
35 40 45

Lys Leu Glu Gly His His His Asp Val Val Ala Cys Asp Phe Ser Pro
50 55 60

Asp Gly Ala Leu Leu Ala Thr Ala Ser Tyr Asp Thr Arg Val Tyr Ile
65 70 75 80

Trp Asp Pro His Asn Gly Asp Ile Leu Met Glu Phe Gly His Leu Phe
85 90 95

Pro Pro Pro Thr Pro Ile Phe Ala Gly Gly Ala Asn Asp Arg Trp Val
100 105 110

Arg Ser Val Ser Phe Ser His Asp Gly Leu His Val Ala Ser Leu Ala
115 120 125

Asp Asp Lys Met Val Arg Phe Trp Arg Ile Asp Glu Asp Tyr Pro Val
130 135 140

Gln Val Ala Pro Leu Ser Asn Gly Leu Cys Cys Ala Phe Ser Thr Asp
145 150 155 160

Gly Ser Val Leu Ala Ala Gly Thr His Asp Gly Ser Val Tyr Phe Trp
165 170 175

Ala Thr Pro Arg Gln Val Pro Ser Leu Gln His Leu Cys Arg Met Ser
180 185 190

Ile Arg Arg Val Met Pro Thr Gln Glu Val Gln Glu Leu Pro Ile Pro
195 200 205

Ser Lys Leu Leu Glu Phe Leu Ser Tyr Arg Ile * Lys Ile Leu Pro
210 215 220

Ser Leu Val Val Gly Thr Asp Arg Ile His Leu Thr Gln Thr Ser Ser
225 230 235 240

Phe Thr Asp Phe Asn Tyr Leu Phe Leu Lys Thr * Lys Ile Tyr Leu
245 250 255

Ile * Tyr Val Leu Val Leu His Phe Asp Gln Leu Lys Leu Leu Lys
260 265 270

Tyr Tyr Leu * Thr Ile Glu Val Phe Leu Asn Ile Ser Asn Ile Asn
275 280 285

Phe Phe Lys Asp Leu Thr Val Lys Asn Ile His Thr Cys Thr Tyr Leu
290 295 300

Asp Ile Ser Cys Tyr Met Leu Asn
305 310

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 257 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
Met Val Leu Cys Val Gln Gly Ser Cys Pro Leu Leu Ala Val Glu Gln
1               5                   10                  15

Ile Gly Arg Arg Pro Leu Trp Ala Gln Ser Leu Glu Leu Pro Gly Pro
            20                  25                  30

Ala Met Gln Pro Leu Pro Thr Gly Ala Phe Pro Glu Glu Val Thr Glu
        35                  40                  45

Glu Thr Pro Val Gln Ala Glu Asn Glu Pro Lys Val Leu Asp Pro Glu
    50                  55                  60

Gly Asp Leu Leu Cys Ile Ala Lys Thr Phe Ser Tyr Leu Arg Glu Ser
65                  70                  75                  80

Gly Trp Tyr Trp Gly Ser Ile Thr Ala Ser Glu Ala Arg Gln His Leu
                85                  90                  95

Gln Lys Met Pro Glu Gly Thr Phe Leu Val Arg Asp Ser Thr His Pro
            100                 105                 110

Ser Tyr Leu Phe Thr Leu Ser Val Lys Thr Thr Arg Gly Pro Thr Asn
        115                 120                 125

Val Arg Ile Glu Tyr Ala Asp Ser Ser Phe Arg Leu Asp Ser Asn Cys
    130                 135                 140

Leu Ser Arg Pro Arg Ile Leu Ala Phe Pro Asp Val Val Ser Leu Val
145                 150                 155                 160

Gln His Tyr Val Ala Ser Cys Ala Ala Asp Thr Arg Ser Asp Ser Pro
                165                 170                 175

Asp Pro Ala Pro Thr Pro Ala Leu Pro Met Ser Lys Gln Asp Ala Pro
            180                 185                 190

Ser Asp Ser Val Leu Pro Ile Pro Val Ala Thr Ala Val His Leu Lys
        195                 200                 205

Leu Val Gln Pro Phe Val Arg Arg Ser Ser Ala Arg Ser Leu Gln His
    210                 215                 220

Leu Cys Arg Leu Val Ile Asn Arg Leu Val Ala Asp Val Asp Cys Leu
225                 230                 235                 240

Pro Leu Pro Arg Arg Met Ala Asp Tyr Leu Arg Gln Tyr Pro Phe Gln
                245                 250                 255

Leu
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 211 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Met Val Ala His Asn Gln Val Ala Ala Asp Asn Ala Val Ser Thr Ala
1               5                   10                  15

Ala Glu Pro Arg Arg Arg Pro Glu Pro Ser Ser Ser Ser Ser Ser Ser
            20                  25                  30

Pro Ala Ala Pro Ala Arg Pro Arg Pro Cys Pro Ala Val Pro Ala Pro
        35                  40                  45

Ala Pro Gly Asp Thr His Phe Arg Thr Phe Arg Ser His Ala Asp Tyr
    50                  55                  60

Arg Arg Ile Thr Arg Ala Ser Ala Leu Leu Asp Ala Cys Gly Phe Tyr
65                  70                  75                  80

Trp Gly Pro Leu Ser Val His Gly Ala His Glu Arg Leu Arg Ala Glu
                85                  90                  95

Pro Val Gly Thr Phe Leu Val Arg Asp Ser Arg Gln Arg Asn Cys Phe
            100                 105                 110

Phe Ala Leu Ser Val Lys Met Ala Ser Gly Pro Thr Ser Ile Arg Val
        115                 120                 125

His Phe Gln Ala Gly Arg Phe His Leu Asp Gly Ser Arg Glu Ser Phe
    130                 135                 140

Asp Cys Leu Phe Glu Leu Leu Glu His Tyr Val Ala Ala Pro Arg Arg
145                 150                 155                 160

Met Leu Gly Ala Pro Leu Arg Gln Arg Arg Val Arg Pro Leu Gln Glu
                165                 170                 175

Leu Cys Arg Gln Arg Ile Val Ala Thr Val Gly Arg Glu Asn Leu Ala
            180                 185                 190

Arg Ile Pro Leu Asn Pro Val Leu Arg Asp Tyr Leu Ser Ser Phe Pro
        195                 200                 205

Phe Gln Ile
    210
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 212 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
Met Val Ala Arg Asn Gln Val Ala Ala Asp Asn Ala Ile Ser Pro Ala
1               5                   10                  15

Ala Glu Pro Arg Arg Arg Ser Glu Pro Ser Ser Ser Ser Ser Ser Ser
            20                  25                  30

Ser Pro Ala Ala Pro Val Arg Pro Arg Pro Cys Pro Ala Val Pro Ala
        35                  40                  45

Pro Ala Pro Gly Asp Thr His Phe Arg Thr Phe Arg Ser His Ser Asp
    50                  55                  60

Tyr Arg Arg Ile Thr Arg Thr Ser Ala Leu Leu Asp Ala Cys Gly Phe
65                  70                  75                  80

Tyr Trp Gly Pro Leu Ser Val His Gly Ala His Glu Arg Leu Arg Ala
                85                  90                  95

Glu Pro Val Gly Thr Phe Leu Val Arg Asp Ser Arg Gln Arg Asn Cys
            100                 105                 110

Phe Phe Ala Leu Ser Val Lys Met Ala Ser Gly Pro Thr Ser Ile Arg
        115                 120                 125

Val His Phe Gln Ala Gly Arg Phe His Leu Asp Gly Ser Arg Glu Thr
    130                 135                 140

Phe Asp Cys Leu Phe Glu Leu Leu Glu His Tyr Val Ala Ala Pro Arg
145                 150                 155                 160

Arg Met Leu Gly Ala Pro Leu Arg Gln Arg Arg Val Arg Pro Leu Gln
                165                 170                 175

Glu Leu Cys Arg Gln Arg Ile Val Ala Ala Val Gly Arg Glu Asn Leu
            180                 185                 190

Ala Arg Ile Pro Leu Asn Pro Val Leu Arg Asp Tyr Leu Ser Ser Phe
        195                 200                 205

Pro Phe Gln Ile
    210
```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 306 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
Ala Leu Ser Pro Ala Ala Thr Leu Thr Ala Trp Pro Ala Asp Ser Ala
1               5                   10                  15
```

Arg Arg Gly Pro Gly Cys Thr Ala Ser Gly Tyr Pro Val Pro Ala Ala
            20   25   30

Arg Ala Pro Ala Ala Gly Asp Gln Trp Val Thr Ala Ala Ala Arg Asp
        35   40   45

Phe Val Ile Arg Pro Pro Gly Ser Gly Glu Lys Glu Pro His Pro Phe
    50   55   60

Ser Leu Cys His His Phe Gly His Pro Ala Gly Leu Val Leu Gly Phe
65   70   75   80

Ala Leu Thr Ser Arg Lys Asp Ala Asn Pro Ser Leu Thr Pro Ala Arg
                85   90   95

Ala Ala Thr Cys Leu Cys Arg Gly Asp Pro Ser Leu Met Thr Leu Arg
            100   105   110

Cys Leu Glu Pro Ser Gly Asn Gly Gly Glu Gly Thr Arg Ser Gln Trp
        115   120   125

Gly Thr Ala Gly Ser Ala Glu Glu Pro Ser Pro Gln Ala Ala Arg Leu
    130   135   140

Ala Lys Ala Leu Arg Glu Leu Gly Gln Thr Gly Trp Tyr Trp Gly Ser
145   150   155   160

Met Thr Val Asn Glu Ala Lys Glu Lys Leu Lys Glu Ala Pro Glu Gly
                165   170   175

Thr Phe Leu Ile Arg Asp Ser Ser His Ser Asp Tyr Leu Leu Thr Ile
            180   185   190

Ser Val Lys Thr Ser Ala Gly Pro Thr Asn Leu Arg Ile Glu Tyr Gln
        195   200   205

Asp Gly Lys Phe Arg Leu Asp Ser Ile Ile Cys Val Lys Ser Lys Leu
    210   215   220

Lys Gln Phe Asp Ser Val Val His Leu Ile Asp Tyr Tyr Val Gln Met
225   230   235   240

Cys Lys Asp Lys Arg Thr Gly Pro Glu Ala Pro Arg Asn Gly Thr Val
                245   250   255

His Leu Tyr Leu Thr Lys Pro Leu Tyr Thr Ser Ala Pro Ser Leu Gln
            260   265   270

His Leu Cys Arg Leu Thr Ile Asn Lys Cys Thr Gly Ala Ile Trp Gly
        275   280   285

Leu Pro Leu Pro Thr Arg Leu Lys Asp Tyr Leu Glu Glu Tyr Lys Phe
    290   295   300

Gln Val
305

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 225 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
Met Val Thr His Ser Lys Phe Pro Ala Ala Gly Met Ser Arg Pro Leu
1               5                   10                  15

Asp Thr Ser Leu Arg Leu Lys Thr Phe Ser Ser Lys Ser Glu Tyr Gln
            20                  25                  30

Leu Val Val Asn Ala Val Arg Lys Leu Gln Glu Ser Gly Phe Tyr Trp
        35                  40                  45

Ser Ala Val Thr Gly Gly Glu Ala Asn Leu Leu Leu Ser Ala Glu Pro
    50                  55                  60

Ala Gly Thr Phe Leu Ile Arg Asp Ser Ser Asp Gln Arg His Phe Phe
65                  70                  75                  80

Ala Leu Ser Val Lys Thr Gln Ser Gly Thr Lys Asn Leu Arg Ile Gln
                85                  90                  95

Cys Glu Gly Gly Ser Phe Ser Leu Gln Ser Asp Pro Arg Ser Thr Gln
            100                 105                 110

Pro Val Pro Arg Phe Asp Cys Val Leu Lys Leu Val Tyr His Tyr Met
        115                 120                 125

Pro Pro Pro Gly Ala Pro Ser Phe Pro Ser Pro Pro Thr Glu Pro Ser
    130                 135                 140

Ser Glu Val Pro Glu Gln Pro Ser Ala Gln Pro Leu Pro Gly Ser Pro
145                 150                 155                 160

Pro Arg Arg Ala Tyr Tyr Ile Tyr Ser Gly Gly Glu Lys Ile Pro Leu
                165                 170                 175

Val Leu Ser Arg Pro Leu Ser Ser Asn Val Ala Thr Leu Gln His Leu
            180                 185                 190

Cys Arg Lys Thr Val Asn Gly His Leu Asp Ser Tyr Glu Lys Val Thr
        195                 200                 205

Gln Leu Pro Gly Pro Ile Arg Glu Phe Leu Asp Gln Tyr Asp Ala Pro
    210                 215                 220

Leu
225
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 225 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
Met Val Thr His Ser Lys Phe Pro Ala Ala Gly Met Ser Arg Pro Leu
1               5                   10                  15

Asp Thr Ser Leu Arg Leu Lys Thr Phe Ser Ser Lys Ser Glu Tyr Gln
            20                  25                  30

Leu Val Val Asn Ala Val Arg Lys Leu Gln Glu Ser Gly Phe Tyr Trp
        35                  40                  45

Ser Ala Val Thr Gly Gly Glu Ala Asn Leu Leu Leu Ser Ala Glu Pro
    50                  55                  60

Ala Gly Thr Phe Leu Ile Arg Asp Ser Ser Asp Gln Arg His Phe Phe
65                  70                  75                  80

Thr Leu Ser Val Lys Thr Gln Ser Gly Thr Lys Asn Leu Arg Ile Gln
                85                  90                  95

Cys Glu Gly Gly Ser Phe Ser Leu Gln Ser Asp Pro Arg Ser Thr Gln
            100                 105                 110

Pro Val Pro Arg Phe Asp Cys Val Leu Lys Leu Val His His Tyr Met
        115                 120                 125

Pro Pro Pro Gly Thr Pro Ser Phe Ser Leu Pro Pro Thr Glu Pro Ser
    130                 135                 140

Ser Glu Val Pro Glu Gln Pro Pro Ala Gln Ala Leu Pro Gly Ser Thr
145                 150                 155                 160

Pro Lys Arg Ala Tyr Tyr Ile Tyr Ser Gly Gly Glu Lys Ile Pro Leu
                165                 170                 175

Val Leu Ser Arg Pro Leu Ser Ser Asn Val Ala Thr Leu Gln His Leu
            180                 185                 190

Cys Arg Lys Thr Val Asn Gly His Leu Asp Ser Tyr Glu Lys Val Thr
        195                 200                 205

Gln Leu Pro Gly Pro Ile Arg Glu Phe Leu Asp Gln Tyr Asp Ala Pro
    210                 215                 220

Leu
225
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 510 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
Leu Tyr Trp Ser Ser Thr Val Val Ala Ala Ala Leu Glu Xaa Xaa Xaa
1               5                   10                  15

Xaa Xaa Gly Cys Xaa Xaa Xaa Glu Xaa Glu Gly Val Arg Ser Ser Pro
            20                  25                  30

Val Val Ser Leu Ser Leu Pro Leu Xaa Arg Ala Arg Met Gly Arg Ala
        35                  40                  45

Glu Leu Leu Glu Gly Lys Met Ser Thr Gln Asp Pro Ser Asp Leu Trp
    50                  55                  60

Ser Arg Ser Asp Gly Glu Ala Glu Leu Leu Gln Asp Leu Gly Trp Tyr
65                  70                  75                  80

His Gly Asn Leu Thr Arg His Ala Ala Glu Ala Leu Leu Leu Ser Asn
                85                  90                  95

Gly Cys Asp Gly Ser Tyr Leu Leu Arg Asp Ser Asn Glu Thr Thr Gly
            100                 105                 110

Leu Tyr Ser Leu Ser Val Arg Ala Lys Asp Ser Val Lys His Phe His
        115                 120                 125

Val Glu Tyr Thr Gly Tyr Ser Phe Lys Phe Gly Phe Asn Glu Phe Ser
    130                 135                 140

Ser Leu Lys Asp Phe Val Lys His Phe Ala Asn Gln Pro Leu Ile Gly
145                 150                 155                 160

Ser Glu Thr Gly Thr Leu Met Val Leu Lys His Pro Tyr Pro Arg Lys
                165                 170                 175

Val Xaa Glu Pro Ser Ile Tyr Glu Ser Val Arg Val His Thr Ala Met
            180                 185                 190

Gln Thr Gly Arg Thr Glu Asp Asp Leu Val Pro Thr Ala Pro Ser Leu
        195                 200                 205

Gly Thr Lys Glu Gly Tyr Leu Thr Lys Gln Gly Gly Leu Val Lys Thr
    210                 215                 220

Trp Lys Thr Arg Trp Phe Thr Leu His Arg Asn Glu Leu Lys Tyr Phe
225                 230                 235                 240

Lys Asp Gln Met Ser Pro Glu Pro Ile Arg Ile Leu Asp Leu Thr Glu
                245                 250                 255

Cys Ser Ala Val Gln Phe Asp Tyr Ser Gln Glu Arg Val Asn Cys Phe
            260                 265                 270

Cys Leu Val Phe Pro Phe Arg Thr Phe Tyr Leu Cys Ala Lys Thr Gly
        275                 280                 285

Val Glu Ala Asp Glu Trp Ile Lys Ile Leu Arg Trp Lys Leu Ser Gln
    290                 295                 300

Ile Arg Lys Gln Leu Asn Gln Gly Glu Ala Arg Ser Asp Leu Gly Arg
305                 310                 315                 320

Ser Ser Leu Asn Arg Ser Phe Leu Pro Arg Asn Ala Leu Ala Gln Glu
                325                 330                 335
```

Gln Val Glu Cys Phe Pro Xaa Arg Cys Asp Leu Xaa Gln Leu Gln Met
340 345 350

Lys Thr Asp Xaa Asp Phe Leu Ser Lys Thr Asn Gln Asn Arg Cys Xaa
355 360 365

Leu Gly Pro Ile Tyr His Val Ala Asp Ser Leu Cys Cys Pro Ser Xaa
370 375 380

Met Leu Pro Xaa Pro Xaa Glu His Xaa Ser Asn His His Xaa Asp Arg
385 390 395 400

Lys Cys Leu Asn His His Ser Xaa Val Cys Ser Leu Leu Glu His Thr
405 410 415

Met Glu Glu Glu Gly Phe Leu Phe Ser Leu Ile Val Val Pro Lys Pro
420 425 430

Ile Asp Thr Ser Cys Leu Glu Ser His Cys Glu Ser Trp Ser Ala Cys
435 440 445

Leu Thr Xaa Arg Leu Cys Tyr Xaa Pro Arg Arg Lys Gln Ile Leu Gly
450 455 460

Gly Leu Asp Asp Xaa Cys Arg Ile Tyr Ile Gln Ile Glu Asn Ile Lys
465 470 475 480

Tyr Phe Gln Gly Arg Gly Phe Phe Phe Xaa Phe Phe Pro Leu Tyr Thr
485 490 495

Lys Lys Lys Lys Lys Lys Leu Glu Gly Gly Pro Tyr Pro Xaa
500 505 510

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2093 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TAAGGTCCAC GTCGCTCCGM AGCCATCACT ACAGKMCCGC GCCGTGGCCT CTGCGGCCCA 60

CAAWCTCCGR GGAGACCTGC ATCAAGATGG AGGTGAGAGT CAAGGCCTTG GTTCACTCTT 120

CCAGCCCGAG TCCAGCCCTG AATGGCGTCC GGAAGGATTT CCACGACCTC CAGTCTGAGA 180

CCACGTGCCA GGAGCAAGCC AATTCACTGA AGAGCTCGGC TTCTCATAAT GGAGACCTGC 240

ATCTTCACCT GGATGAACAT GTGCCTGTCG TTATTGGACT TATGCCTCAG GACTACATTC 300

AGTATACTGT GCCTTTAGAT GAGGGGATGT ATCCTTTGGA AGGATCACGG AGCTATTGTC 360

TGGACAGCTC TTCTCCCATG GAAGTCTCTG CGGTTCCTCC TCAAGTGGGA GGGCGCGCTT 420

TCCCCGAGGA TGAGAGTCAG GTAGACCAGG ACCTAGTTGT CGCCCCAGAG ATCTTCGTGG 480

```
ATCAGTCCGT GAATGGCTTG TTGATTGGCA CCACGGGAGT CATGTTGCAG AGCCCGAGAG    540
CGGGTCACGA TGATGTCCCT CCACTCTCAC CATTGCTACC TCCAATGCAG AATAATCAAA    600
TCCAAAGGAA CTTCAGTGGA CTCACTGGCA CAGAAGCCCA CGTGGCTGAA AGTATGCGCT    660
GTCATTTGAA TTTTGATCCG AACTCTGCTC CTGGGGTTGC AAGAGTTTAT GACTCAGTGC    720
AAAGTAGTGG TCCCATGGTT GTGACAAGCC TTACAGAGGA GCTGAAAAAA CTTGCAAAGC    780
AAGGATGGTA CTGGGGACCA ATCACACGTT GGGAGGCAGA AGGGAAGCTA GCAAACGTGC    840
CAGATGGTTC TTTTCTTGTT CGGGACAGTT CTGACGACCG TTACCTTTTA AGCTTGAGCT    900
TTCGCTCCCA TGGTAAAACA CTTCACACTA GAATTGAGCA CTCAAATGGT AGGTTTAGCT    960
TTTATGAACA GCCAGATGTG GAAAGGACAT ACTCCATAGT TGATCTAATT GAGCATTCCA   1020
TCCAGGGACT CGAAAATGGA GCTTTTTGTT ATTCAAGGTC TCGGCTGCCT GGATCTGCAA   1080
CTTACCCCGT CAGACTGACC AACCCAGTGT CCCGGTTCAT GCAGGTGCGC TCGTTGCAGT   1140
ACCTGTGTCG TTTTGTTATA CGTCAGTATA CCAGAATAGA CTTAATTCAG AAACTGCCTT   1200
TGCCAAACAA AATGAAGGAT TATTTACAGG AGAAGCACTA CTGAAAGATT GAGAACCCTG   1260
CATCTTGCAC TTTGGGAATA AGAACAAGAG ATTGAAATAC AGTTTACAAA CTTTCATTGC   1320
CATCAAAATC TTTTGCTGCC ATAACTATTT CAGTTTTATG TGTAAAAGAG TCATCAGTTT   1380
GTTTAGGGGT GGGGAAGTGT CAGCAAGGTG TCTTGGGTTT ATTTTGGTTC TTTAAAAAAG   1440
GGAAGTCTTG AAGTTTTAGA RGTGTTGAAT TATGTTTCAT CAATGTGCAG AATAATCACA   1500
ATGTGAATTA TCAAATTCTC CTCAATGCCC CCCCCGCCCA KTCCTTTGCT GCTATCCACT   1560
GTGATTTTTA TGCATTAAAA GCMCATTTCA TGTKTTTTCA ACCCTAAGTA AAGTTGAATG   1620
AAACTTAACR GAATGGAAAT TGCTATTTCT TTTTAAATGG YCCATTTTCC AAAAMARGTG   1680
TTGAATAAMC AWMCCTGTKT GAATAAAACM MGRAWTTWMM WWTARCAMYG BAGRTGRGTT   1740
TTTAATCTYY TAMYTTDAAA AGATTTATTT AGAATYGKKA ATTGACMTAA TATTGGGTWA   1800
TBGGRMCGGR GATCTGSAAC ATATKYTTTA ACAACAWTTT WTTKKCYTTA ATKKDTTTYY   1860
AARGKTGGBC TTATTWHTTT GGBKBBSVAA AGKWBVAHTT CTCYGTYSCC YTCGTTTTCA   1920
TCTTCTAGTT TGTGNTATTT TAATAAATGG CCTTACATTA AAAAATTGTA AAGAAATGTA   1980
TACCACCAAT TTAGAAATTG TTGCCTTTTC TGTAATTAAA CTCGGGTACA AATNGGCATA   2040
ACATGAAAAC CTATGGAACT AGAATTATTA TTAAAGAAAT ATTAGATGAT CAT          2093
```

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1748 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ATGGAGGCCG GAGAGGAACC GCTGCTGCTG GCCGAACTCA AGCCCGGGCG CCCCCACCAG 60
TTTGATTGGA AGTCCAGCTG TGAAACCTGG AGCGTGGCCT TCTCGCCAGA CGGTTCCTGG 120
TTCGCCTGGT CTCAAGGACA CTGCGTGGTC AAGCTGGTCC CCTGGCCCTT AGAGGAACAG 180
TTCATCCCTA AAGGATTCGA AGCCAAGAGC CGAAGCAGCA AGAATGACCC AAAAGGACGG 240
GGCAGTCTGA AGGAGAAGAC GCTGGACTGT GGCCAGATTG TGTGGGGGCT GGCCTTCAGC 300
CCATGGCCCT CTCCACCCAG CAGGAAACTC TGGGCACGTC ACCATCCCCA GGCGCCTGAT 360
GTTTCTTGCC TGATCCTGGC CACAGGTCTC AACGATGGGC AGATCAAGAT TTGGGAGGTA 420
CAGACAGGCC TCCTGCTTCT GAATCTTTCT GGCCACCAAG ACGTCGTGAG AGATCTGAGC 480
TTCACGCCCA GCGGCAGTTT GATTTTGGTC TCTGCATCCC GGGATAAGAC ACTTCGAATT 540
TGGGACCTGA ATAAGCACGG TAAGCAGATC CAGGTGTTAT CCGGCCATCT GCAGTGGGTT 600
TACTGCTGCT CCATCTCCCC TGACTGTAGC ATGCTGTGCT CTGCAGCTGG GGAGAAGTCG 660
GTCTTTCTGT GGAGCATGCG GTCCTACACA CTAATCCGGA AACTAGAAGG CCACCAAAGC 720
AGTGTTGTCT CCTGTGATTT CTCTCCTGAT TCAGCCTTGC TTGTCACAGC TTCGTATGAC 780
ACCAGTGTGA TTATGTGGGA CCCCTACACC GGCGAGAGGC TGAGGTCACT TCATCACACA 840
CAGCTTGAAC CCACCATGGA TGACAGTGAC GTCCACATGA GCTCCCTGAG GTCCGTGTGC 900
TTCTCACCTG AAGGCTTGTA TCTCGCTACG GTGGCAGATG ACAGRCTGCT CAGGATCTGG 960
GCTCTGGAAC TGAAAGCTCC GGTTGCCTTT GCTCCGATGA CCAATGGTCT TTGCTGCACA 1020
TTTTTYCCAC AYGGTGGAAT YATTGCCACA GGGACAAGAG ATGGCCACGT CCAGTTCTGG 1080
ACAGCTCCTA GGGTCCTGTC CTCACTGAAG CACTTATGCC GGAAAGCCCT TCGAAGTTTC 1140
CTAACAACTT ACCAAGTCCT AGCACTGCCA ATCCCCAAGA AAATGAAAGA GTTCCTCACA 1200
TACAGGACTT TTTAAGCAAC ACCACATCTT GTGCTTCTTT GTAGCAGGGT AAATCGTCCT 1260
GTCAAAGGGA GTTGCTGGAA TAATGGGCCA AACATCTGGT CTTGCATTGA AATAGCATTT 1320
CTTTGGGATT GTGAATAGAA TGTAGCAAAA CCAGATTCCA GTGTACTAGT CATGGRTCTT 1380
TCTCTCCCTG GGCATGTGGA AAGTCAGTCT TAGGAGGGAA GGAGATTCCA CTTGKCACGG 1440
GCAACAGAGC CYTTACGTTT AAATTTTTCA GTCCAGTTAT KGAACAGCAA GTGTTTGAAM 1500
TCTTTCTGGY TTGTTTTKGA WTTCAAAGTG GCAGTTACTG RWKGTTGTTT TTGGATTTAT 1560
GGCAACYAAG TTAGGGCCTC CAGNGGTTNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN 1620
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNT HNABNVNRNN NRTNNNNRMA TNNNNNNNNN 1680
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN 1740
NNNNNNNN 1748

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2198 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
GGCGGTGGTG ATGGCGGCAG GCGCTCGGAC AGCTCCGCTT GAGCTGAGCT CGGAGAGATC     60
CGTCCAGAAA GTGCCCAGAA GAAACTTCCT CTTAGAAAAG CTGAAAACAC AATATTTATA    120
ACACTGGAAA TTGTAAAGAA TTTGTTTAAA ATGGCTGAAA ACAATAGTAA AAATGTAGAT    180
GTACGGCCTA AAACAAGTCG GAGTCGAAGT GCTGACAGGA AGGATGGTTA TGTGTGGAGT    240
GGAAAGAAGT TGTCTTGGTC CAAAAAGAGT GAGAGTTGTT CTGAATCTGA AGCCAAGAAA    300
GGGCAGCTTA GCTGTTCGTC CATTGAGTTG GACTTAGATC ATTCCTGTGG GCATAGATTT    360
TTAGGCCGAT CCCTTAAACA GAAACTGCAA GATGCGGTGG GGCAGTGTTT TCCAATAAAG    420
AATTGTAGTG GCCGACACTC TCCAGGGCTT CCATCTAAAA GAAAGATTCA TATCAGTGAA    480
CTCATGTTAG ATAMGTGYSC YTTCCCACCT CGCTCAGATT TAGCCTTTAG GTGGCATTTT    540
ATTAAACGAC ACACTGTTCC TATGAGTCCC AACTCAGATG AATGGGTGAG TGCAGACCTG    600
TCTGAGAGGA AACTGAGAGA TGCTCAGCTG AAACGAAGAA ACACAGAAGA TGACATACCC    660
TGTTTCTCAC ATACCAATGG CCAGCCTTGT GTCATAACTG CCAACAGTGC TTCGTGTACA    720
GGTGGTCACA TAACTGGTTC TATGATGAAC TTGGTCACAA ACAACAGCAT AGAAGACAGT    780
GACATGGATT CAGAGGATGA AATTATAACG CTGTGCACAA GCTCCAGAAA AAGGAATAAG    840
CCCAGGTGGG AAATGGAAGA GGAGATCCTG CAGTTGGAGG CACCTCCTAA GTTCCACACC    900
CAGATCGACT ACGTCCACTG CCTTGTTCCA GACCTCCTTC AGATCAGTAA CAATCCGTGC    960
TACTGGGGTG TCATGGACAA ATATGCAGCC GAAGCTCTGC TGGAAGGAAA GCCAGAGGGC   1020
ACCTTTTTAC TTCGAGATTC AGCGCAGGAA GATTATTTAT TCTCTGTTAG TTTTAGACGC   1080
TACAGTCGTT CTCTTCATGC TAGAATTGAG CAGTGGAATC ATAACTTTAG CTTTGATGCC   1140
CATGATCCTT GTGTCTTCCA TTCTCCTGAT ATTACTGGGC TCCTGGAACA CTATAAGGAC   1200
CCCAGTGCCT GTATGTTCTT TGAGCCGCTC TTGTCCACTC CCTTAATCCG GACGTTCCCC   1260
TTTTCCTTGC AGCATATTTG CAGAACGGTT ATTTGTAATT GTACGACTTA CGATGGCATC   1320
GATGCCCTTC CCATTCCTTC GCCTATGAAA TTGTATCTGA AGGAATACCA TTATAAATCA   1380
AAAGTTAGGT TACTCAGGAT TGATGTGCCA GAGCAGCAGT GATGCGGAGA GGTTAGAATG   1440
```

TCKACCGGAG CTTTYGTTCC CTTTAGTGAG GGTTAATTTC GAGCTTGGCG TAATCATGGT 1500

CATAGCTGTT TCCTGTGTGA AATYGTYATC CGCTCACAAT TCCACACAAC ATACGAGCCG 1560

GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG CTAACTCACA TTAATTGSGT 1620

YGCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG CCASCTGCAT TAMTGAATCN 1680

GCCAACKCGC NGGGANAGCG GTTNGCNTAT TGGGCGCTCT TCACTTCNTC GCTCACTGAN 1740

TCNCTNCCTC GGTCNTTCGN TGCTGCTACN GTNTCCCCCA TCCAAGCGTT ATACGCTATC 1800

CNCAGAACTG GGAAANNCNG AANACNNTNA CAAAGCTCAN TGCTANCGTA NACGCCNTGC 1860

NGGCTTTTCC TCGTCCCCCN ACACNCTAAA CAGCCCTCGA GTGCAACCNC GATATANATN 1920

TCTTCCCTNA ACCCCTGCCT CTGTCNCCGC CTNCGACTTC GCTTCCNNGG ATTGCTTTCN 1980

CCCCGTAGTC NGTCNTAGTG NGCNGCGCCT TCCACCCTTC NACCNCTACG TANNNNNANN 2040

CNCCAAANCC NCCNCCCCTC NGATAAAAAG TNAGNGCCTT NANNNCCNNG ATAAAAATGG 2100

TCCCNTACTT TCCAATGTCT NCCNCCCGGC TNTTCTNGCC ACCCAANTNA NNTTTCCGGN 2160

ACTGNATCCG GTGCTANCNT CCTGTTTCTC CTCCCNCC 2198

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2254 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CTCGGGCCGG GATGGATCCG CCGGGAAGAG GAAGACAAGC GGAGCGTTGA GCCCCTGCGC 60

ACGGTGCCCC GCGCGTAGTG GGAGCTTACT CGCAGTAGCT CTCGCTCTTC TAATCAATGG 120

ATAAAGTGGG GAAAATGTGG AACAACTTAA AATACAGATG CCAGAATCTC TTCAGCCACG 180

AGGGAGGAAG CCGTAATGAG AACGTGGAGA TGAACCCCAA CAGATGTCCG TCTGTCAAAG 240

AGAAAAGCAT CAGTCTGGGA GAGGCAGCTC CCCAGCAAGA GAGCAGTCCC TTAAGAGAAA 300

ATGTTGCCTT ACAGCTGGGA CTGAGCCCTT CCAAGACCTT TTCCAGGCGG AACCAAAACT 360

GTGCCGCAGA GATCCCTCAA GTGGTTGAAA TCAGCATCGA GAAAGACAGT GACTCGGGTG 420

CCACCCCAGG AACGAGGCTT GCACGGAGAG ACTCCTACTC GCGGCACGCC CCGTGGGGAG 480

GAAAGAAGAA ACATTCCTGT TCCACAAAGA CCCAGAGTTC ATTGGATACC GAGAAAAAGT 540

TTGGTAGAAC TCGAAGCGGC CTTCAGAGGC GAGAGCGGCG CTATGGAGTC AGCTCCATGC 600

AGGACATGGA CAGCGTTTCT AGCCGCGCGG TCGGGAGCCG CTCCCTGAGG CAGAGGCTCC 660

AGGACACGGT GGGTTTGTGT TTTCCCATGA GAACTTACAG CAAGCAGTCA AAGCCACTCT 720

TTTCCAATAA AAGAAAAATM CATCTYTCTG AATTAATGCT KGAGAAATGC CCTTTTCCTG 780
CTGGCTCRGA TTTAGCMCAA AAGTGGCATT TGATTAAACA GCATACAGCT CCTGTGAGCC 840
CACATTCAAC ATTTTTTGAT ACRTTTGATC CATCTTTGGT TTCTACAGAA GATGAAGAAG 900
ATAGGCTTAG AGAGAGAAGG CGGCTTAGTA TTGAAGAAGG GGTTGATCCC CCTCCCAATG 960
CACAAATACA TACATTTGAA GCTACTGCAC AGGTTAATCC ATTATTTAAA CTGGGACCAA 1020
AATTAGCTCC TGGAATGACT GAAATAAGTG GGGACAGTTC TGCAATTCCA CAAGCTAATT 1080
GTGACTCGGA AGAGGATACA ACCACCCTGT GTTTGCAGTC ACGGAGGCAG AAGCAGCGTC 1140
AGATATCTGG AGACAGCCAT ACSCATGTTA GCAGACAGGG AGCTTGGAAA GTCCACACAC 1200
AGATTGATTA CATACACTGC CTCGTGCCTG ATTTGCTTCA AATTACAGGG AATCCCTGTT 1260
ACTGGGGAGT GATGGACCGT TATGAAGCAG AAGCCCTCTC CGAAGGGAAA CCKGAAGGCA 1320
CGTTCTTGCT CAGGGACTCT GCACAGGAGG ACTACCTCTT CTCTGTGAGT TCCGCCGCTA 1380
CAACAGGATC TCTGCACGCC CGGATCGAGC AGTGGAACCA CAACTTCAGC TTCGATGCCC 1440
ATGACCCCTG CGTGTTTCAY TCCTCCACTG TCACGGGGCT TCTCGAACAC TATAAAGAYC 1500
CCAGTTCKTG CATGTTTTTT GAACCGTTGC TAACGATATC ACTSAATAGR ACTTTCCCTT 1560
TCAGCCTGCA GTATATCTGC CGCGCAGTGA TCTGCAGATG CACTACGTAT GATGGGATTG 1620
ACGGGCTCCC GCTACCGTCG ATGTTACAGG ATTTTTTAAA AGAGTATCAT TATAAACAAA 1680
AAGTTAGAGT TCGCTGGTTG GAACGAGAAC CAGTCAAGGC AAAGTAAACT CTCCGGTCCC 1740
CAAAGGGTGT TAACTAGGTC CGCTTTCATG TGCATCAGAC AGTACACCTA TAGCAAGCAC 1800
ACGTAGCAGT GTTAGGCTTT TTCATACAGT ATGTAAGCTT AGTGTTAGTA TCTGTCAGAT 1860
GCTACCTGCT GTTACTTATT CAGATAAACA TGGTGCCTAT TGGAACAATA GCGGATAGAG 1920
CTACAGGTGT TCAGTAAGAC TACAAAAACA TTTTGCCTAT TTCGCTAACA GTTTGGTTTT 1980
TAATGGCTGT GGTATTTGAG TGAGGCAAYY CTGGGGCATT TGTTATGAAG AATTCTATTT 2040
CTTACTGAAG AACAAATWAT TAATATTGGA TGAGTATTTC AACAGTGTGA CTAATGTTTG 2100
AAATTATTTT TTCCTAAGAG TTTTTCCWAT AACCTTCCMA AAGTCGTGAT GTTTGTAGTT 2160
ACCATAATCC AGCTTTGRAG TCCMAAARGA TTAAAGRCYG CCTCCCTTTG RAAAATGCCA 2220
TTTCYKGCCC CAAGGCCTAG TGCCGTCCCT NCGG 2254

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2206 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GGAGCGCGGC CTGGAGACTA ACAGCTGCTC GGAAGAGGAG CTCAGCAGCC CGGGTCGCGG 60
AGGAGGAGGG GGCGGCCGGC TTCTGCTGCA GCCCCCAGGC CCTGAATTAC CTCCGGTGCC 120
CTTCCCGCTG CAGGACTTGG TCCCTCTGGG GCGCCTGAGT AGAGGGGAGC AGCAGCAGCA 180
GCAGCAGCAG CAACCTCCCC CGCCCCCGCC TCCTCCCGGG CCCCTCCGGC CACTCGCGGG 240
TCCTTCTCGG AAGGGCTCCT TCAAAATCCG CCTCAGTCGC CTCTTTCGCA CCAAGAGCTG 300
CAACGGTGGC TCCGGCGGTG GGGATGGGAC CGGCAAGAGG CCTTCTGGAG AGCTGGCTGC 360
TTCAGCTGCG AGCCTGACAG ACATGGGAGG CTCTGCGGGC CGGGAGCTGG ACGCGGGGAG 420
GAAACCCAAG TTGACAAGAA CTCAAAGTGC CTTTTCTCCG GTCTCCTTCA GCCCCCTGTT 480
CACAGGTGAA ACTGTGTCGC TTGTGGATGT GGACATTTCT CAGCGGGGCC TGACCTCTCC 540
ACACCCTCCA ACTCCCCCTC CTCCTCCGAG AAGAAGCCTC AGCCTCCTAG ATGATATCAG 600
TGGGACGCTG CCTACATCTG TCCTTGTGGC TCCGATGGGG TCTTCCTTGC AGTCTTTCCC 660
CCTACCTCCG CCTCCTCCAC CCCATGCCCC AGATGCATTT CCCCGGATTG CTCCCATCCG 720
AGCAGCTGAA TCCCTGCACA GCCAACCCCC ACAGCACCTC CAGTGTCCCC TCTACCGGCC 780
TGACTCGAGC AGCTTTGCAG CCAGCCTTCG AGAGTTGGAG AAGTGTGGTT GGTATTGGGG 840
GCCAATGAAT TGGGAAGATG CAGAGATGAA GCTGAAAGGG AAACCAGATG GTTCTTTCCT 900
GGTACGAGAC AGTTCTGATC CTCGTTACAT CCTGAGCCTC AGTTTCCGAT CACAGGGTAT 960
CACCCACCAC ACTAGAATGG AGCACTACAG AGGAACCTTC AGCCTGTGGT GTCATCCCAA 1020
GTTTGAGGAC CGCTGTCAAT CTGTTGTAGA GTTTATTAAG AGAGCCATTA TGCACTCCAA 1080
GAATGGAAAG TTTCTCTATT TCTTAAGATC CAGGGTTCCA GGACTGCCAC CAACTCCTGT 1140
CCAGCTGCTC TATCCAGTGT CCCGATTCAG CAATGTCAAA TCCCTCCAGC ACCTTTGCAG 1200
ATTCCGGATA CGACAGCTCG TCAGGATAGA TCACATCCCA GATCTCCCAC TGCCTAAACC 1260
TCTGATCTCT TATATCCGAA AGTTCTACTA CTATGATCCT CAGGAAGAGG TATACCTGTC 1320
TCTAAAGGAA GCGCAGCTCA TTTCCAAACA GAAGCAAGAG GTGGAACCCT CCACGTAGCG 1380
AGGGGCTCCC TGCTGGTCAC CACCAAGGGC ATTTGGTTGC CAAGCTCCAG CTTTGAAGAA 1440
CCAAATTAAG CTACCATGAA AAGAAGAGGA AAAGTGAGGG AACAGGAAGG TTGGGATTCT 1500
CTGTGCAGAG ACTTTGGTTC CCCACGCAGC CCTGGGGCTT GGAAGAAGCA CATGACCGTA 1560
CTCTGCGTGG GGCTCCACCT CACACCCACC CCTGGGCATC TTAGGACTGG AGGGGCTCCT 1620
TGGAAAACTG GAAGAAGTCT CAACACTGTT TCTTTTTCAA AAAAAAAAAA AAAAAAGATG 1680
CGGCCGCAAG CTTATTCCCT TTAGTGAGGG TTAATTTTAG CTTGGCACTG GCCGTCGTTT 1740
TACAACGTCG TGACTGGGAA AACCCTGGCG TTACCCAACT TAATCGCCTT GCAGCACATC 1800

CCCCTTTCGC CAGCTGGCGT AATAGCGAAG AGGCCCGCAC CGATCGCCCT TCCCAACAGT 1860

TGCGCAGCCT GAATGGCGAA TGGGACGCGC CCTGTAGCGG CGCATTAACG CGCGGCGGGT 1920

GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG CCCTACGCCC GCTCCTTTCG 1980

CTTTCTTCCC TTCCTTTCTC GCCACGTTCG CCGGCTTTCC CCGTCAACTC TAAATCGGGG 2040

GCTCCCTTTA GGTTCCGATT TANTGCTTTA CGCACTCNAC CCCAAAACTT GATTAGGTGA 2100

TGTCACTTAT GGCACNCCTG ATAACGTTTC CCCTTACTTT GATCACTTCT TTATATGATC 2160

TTTCCAATGA AACATCACCT ACTCGTCATC TTTATTTAAA GATTTG 2206

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1390 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CGGACGCGTG GGTTTGGCTG TGAATATTCT ATTTGCTTGC AGTATCTGTT TCTCTTCCTA 60

GGCTCAAGTT GGTGACCCAA GCCTATTGTA AACAAGTGAT TATCTCANNG GGAGATGCCA 120

ATGGAGTAAC AATTTGTTAA CCTTACGTTT TCTGTCTGTA TATTTTTTTA AAAATCTGGT 180

AGTTTCTGGA AAAAAAAGAG AAGGGGGTTT GTAGTACTTA ACCCTATTTA TTKSCRYRWG 240

TTTTAGTTAA TTAGTTTTTG GAATAAATGG ATTTCAGTAT AGCTTTGTGG TTAAATTGCA 300

TTGCCTTTAT TTTATGTTTA GGCTTATTTT TAAATTAACA TTTAACAGAA ACATTTGAAA 360

TAGAATTTGC ATGTCTGCCT TAATTAACTT AAAGACTGAT TTTAATCTGA CTATGACACT 420

GAGCATATTC TTTAAATTAC TCATAATTTA TAATGCTTAA TATAATCTTA ATTAAATTTA 480

GCAGTTTTAG TATAAGATGT GCCATTTTGT CCTCTGTATG TCTGAATGAA GCTATAACAT 540

TTGCCTTTTT ATTGCAGGTT TTCCTTTGGA ATATGGATAA ATACACCATG ATACGGAAAC 600

TAGAAGGACA TCACCATGAT GTGGTAGCTT GTGACTTTTC TCCTGATGGA GCATTACTGG 660

CTACTGCATC TTATGATACT CGAGTATATA TCTGGGATCC ACATAATGGA GACATTCTGA 720

TGGAATTTGG GCACCTGTTT CCCCCACCTA CTCCAATATT TGCTGGAGGA GCAAATGACC 780

GGTGGGTACG ATCTGTATCT TTTAGCCATG ATGGACTGCA TGTTGCAAGC CTTGCTGATG 840

ATAAAATGGT GAGGTTCTGG AGAATTGATG AGGATTATCC AGTGCAAGTT GCACCTTTGA 900

GCAATGGTCT TTGCTGTGCC TTCTCTACTG ATGGCAGTGT TTTAGCTGCT GGGACACATG 960

ACGGAAGTGT GTATTTTTGG GCCACTCCAC GGCAGGTCCC TAGCCTGCAA CATTTATGTC 1020

GCATGTCAAT CCGAAGAGTG ATGCCCACCC AAGAAGTTCA GGAGCTGCCG ATTCCTTCCA 1080

AGCTTTTGGA GTTTCTCTCG TATCGTATTT AGAAGATTCT GCCTTCCCTA GTAGTAGGGA 1140

CTGACAGAAT ACACTTAACA CAAACCTCAA GCTTTACTGA CTTCAATTAT CTGTTTTTAA 1200

AGACGTAGAA GATTTATTTA ATTTGATATG TTCTTGTACT GCATTTTGAT CAGTTGARGC 1260

TTTTAAAATA TTATTTATAG ACAATAGAAG TATTTCTGAA CATATCAAAT ATAAATTTTT 1320

TTAAAGATCT AACTGTGAAA AACATACATA CCTGTACATA TTTAGATATA AGCTGCTATA 1380

TGTTGAATGG 1390